Multi-object segmentation using shape particles

Marleen de Bruijne and Mads Nielsen

IT University of Copenhagen, Denmark
marleen@itu.dk

Abstract. Deformable template models, in which a shape model and its corresponding appearance model are deformed to optimally fit an object in the image, have proven successful in many medical image segmentation tasks. In some applications, the number of objects in an image is not known a priori. In that case not only the most clearly visible object must be extracted, but the full collection of objects present in the image.
We propose a stochastic optimization algorithm that optimizes a distribution of shape particles so that the overall distribution explains as much of the image as possible. Possible spatial interrelationships between objects are modelled and used to steer the evolution of the particle set by generating new shape hypotheses that are consistent with the shapes currently observed.
The method is evaluated on rib segmentation in chest X-rays.

1 Introduction

Statistical shape models are widely applied in image segmentation [1–3], and are powerful tools especially in the case of missing or locally ambiguous boundary evidence. Most approaches perform a local optimization after the shape model has been initialized on the average position in the image. Alternatively, the best result of a set of local optimizations with different initializations can be selected. In the case of multiple objects, one would typically construct a combined model of all objects and optimize all shapes simultaneously in the image.

In some cases, modelling all objects jointly is not desirable. There may be not enough training data available to construct a sufficiently flexible and accurate model, rotation or scaling of one object with respect to another may introduce unwanted non-linearities in the model, and optimization in a high dimensional space is computationally more expensive. Moreover, if the number of objects present in the image is unknown it is impossible to define corresponding points in all images.

This paper presents a solution to the problem of segmenting an unknown number of (similar) objects. The segmentation is represented by a distribution of shape 'particles' that evolves under the influence of image terms and interaction between neighboring shapes. The particle cloud evolution is similar to Monte Carlo methods known as 'Condensation', 'particle filtering', or 'factored sampling', which have been applied to object localization and tracking [4–6]. However, the definition of the image term is different in this case in which multiple objects are modelled with the same particle distribution.

In a previous paper, we proposed the use of particle filtering to optimize shape-classification templates on a probability map obtained from pixel classification. Shape particles are weighted by their likelihood and the particle distribution is evolved using weighted resampling and a small amount of random perturbation in each iteration. In this way, particles representing unlikely shapes vanish while successful particles multiply. The initial sparse sampling evolves into a δ-peak at the maximum likelihood solution [7].

In the current paper we seek to optimize not just one object, but the entire shape distribution. The segmentation is represented as the maximum likelihood (soft) classification of the distribution of shape particles. The weights of the particles are adjusted so that the classification obtained from the shape set approximates the observed pixel classification as best as possible. The shape set thus evolves into the maximum likelihood shape collection.

Spatial consistency of the total segmentation can be enforced by neighbor interactions between particles. In the particle diffusion step each particle is allowed to produce hypotheses for neighboring shapes on basis of its own shape and position and a learned conditional shape model. This is especially useful in regular shape patterns such as the spine or the rib cage. If one vertebra or rib is found in the image, there is a high probability that a second vertebra or rib is present with approximately the same shape but a few centimeters higher or lower.

We applied this method to segmenting the ribs in the lung fields in chest radiographs. Rib segmentations are used for instance as a frame of reference for localizing abnormalities such as lung nodules, and to eliminate false positives in abnormality detection that frequently occur at crossings of posterior and anterior parts of the ribs. Classical approaches which fit geometrical models to edges in the image may miss some ribs and detect other ribs twice [8]. Loog [9] combined gray value features and contextual features in an iterative classification scheme, thus learning an implicit model of local rib structure. Although this produced significantly smoother and more accurate results than pixel classification based on intensity features alone, in some cases ribs were completely missed or the clavicles mistaken for ribs. The fact that consecutive ribs often have similar shapes and are regularly spaced calls for a global shape model describing the relations between different ribs, but construction and optimization of a global shape model is problematic since the number of ribs visible in the lung fields can vary. Ramachandran et al [10] showed that a pre classification of training images by the number of visible spaces between the ribs significantly improves the success of active shape model (ASM) segmentation.

In this work we do not model the full rib cage, but instead model separate ribs and fit those to the image in a consistent pattern using a model of spatial interrelationships between neighboring rib shapes. Neighbor relations between successive ribs in the same lung field as well as between the ribs at the same height in the opposite lung field are modelled.

Section 2 explains the main ideas behind the estimation of a maximum likelihood shape collection whereas Section 3 explains further how these ideas can

be brought into practice. Section 4 gives more details on all ingredients required for multiple object segmentation and the algorithm proposed. Specific choices for rib segmentation and experiments on chest X-rays are described in Section 5. Sections 7 and 8 provide a discussion and conclusions.

2 Image explanation as maximum likelihood

The standard approaches of fitting a shape model to an image can be written in Bayesian formulation as

$$p(S|I) \propto p(I|S)p(S)$$

where the shape S is searched for in the image I either as maximum a posteriori (MAP) estimate maximizing $p(S|I)$ or as maximum likelihood estimate (ML) maximizing $p(I|S)$. In both cases the maximum likelihood term may be evaluated as an object fit assuming an appropriate spatially independent noise model, maximizing

$$\log p(I|S) \propto \int_{\Omega(S)} U(I(x), S(x,\theta))dx \tag{1}$$

where x are the image coordinates, θ contains the shape model parameters to be optimized, U is the local log-likelihood function, and $\Omega(S)$ is the spatial domain of $S(\cdot,\theta)$. In the following we will generalize this to a collection of shapes.

Let $\mathcal{S}$ denote a collection of N shape instances:

$$\mathcal{S} = \{S_1, S_2, \dots, S_N\}$$

Now one straightforward generalization of the likelihood term (Eq. 1) is the sum of the individual terms:

$$\sum_i \int_{\Omega(S_i)} U(I(x), S_i(x,\theta))dx$$

However, this sum of model fits of individual shapes is not the same as the likelihood of the collection of shapes. Optimization of θ with respect to this sum would result in all shapes fitting to the one object with the strongest image evidence. This is due to a simplification made in single-shape modelling that the integration area is only over the shape model. For a proper ML-estimation, all data must be modelled, and the integration domain is the full image domain. The collection likelihood then reads:

$$\log p(I|\mathcal{S}) \propto \int_{\Omega(\mathcal{S})} U(I(x), \mathcal{S}(x,\theta))dx$$

Hence, in every position in the image, it is necessary to take all (overlapping) shapes into account.

3 Computational approach

Let us assume, that the data $I(x)$ takes values in a discrete set $C = \{c_1, c_2, \ldots, c_K\}$ where c_i can be integer pixel values or, as in the example below, pixel classes. Let us now for simplicity assume that in a given pixel, the pixel class due to the individual shapes is given deterministically and independent of the other shapes in the collection, so that

$$p(c_j|\mathcal{S}) = \sum_i p(c_j|S_i)p(S_i) = \sum_i \delta(c(S_i(x)), c_j)p(S_i) \tag{2}$$

In the simple case where all the overlapping shapes are equally probable, this is simply the fraction of overlapping shapes that vote for the class c_i.

Let us now represent the shape collection $\mathcal{S}$ by a weighted set of shape instances S_i with shape parameters θ_i. The weights α_i denote the relative probability of these shape instances. In order to estimate the maximum likelihood shape collection, we must simultaneously optimize over $S_i = S(\theta_i)$ and α_i.

This optimization can be achieved by particle filtering iterating over

$$\begin{array}{l} \alpha = \arg\max_\alpha \ p(I|\theta, \alpha) \\ \theta = \text{sampling} \ \ p(S|\mathcal{S}) \end{array} \tag{3}$$

The first equation may be solved for analytically or obtained by stochastic optimization. If an infinite number of particles $S(\theta_i)$ were available, $\mathcal{S}(\theta, \alpha)$ is the maximum likelihood shape collection after this first optimization. To make the optimization efficient, we start out with a sparse sampling in $p(S)$ and condense this distribution around likely shape collections by the sampling step in Eq. 3. This sampling may be realized by first sampling in $p(S_i) = \alpha_i$ and then in $p(S|S_i)$. The distribution $p(S|S_i)$ represents the belief in S being a true shape in the image when the shape S_i has been observed in the collection.

If the variance of $p(S|S_i)$ decreases to zero in successive iterations the choice of distribution does not influence the point of convergence, as long as $p(S_1|S_2) = p(S_2|S_1)$ and the distribution can explore the full solution. The algorithm is guaranteed to converge to the maximum likelihood solution for a collection of shapes. The proof is analogous to the proof for the individual shape fitting by particle filtering [7].

However, the rate of convergence can be improved by choosing $p(S|S_i)$ so as to explore the solution space most effectively. Here, knowledge of spatial relationships between different objects can be exploited by letting a selected particle S_i produce a plausible hypothesis for a neighboring shape. $p(S|S_i)$ could then be a mixture of densities describing both the uncertainty in the observation S_i and the conditional densities $p(Sn|S_i)$ of all possible neighbors Sn given the observation S_i.

If the variance in $p(S|S_i)$ does not vanish during iteration, the distribution converges to the maximum likelihood shape collection convolved with $p(S|S_i)$.

4 Implementation

This section describes in more detail all ingredients needed to perform multi-object shape model segmentation, viz. a shape model, an image appearance model, possible neighbor relations, and an optimization algorithm.

4.1 Shape model

To constrain the shape of possible solutions, any kind of shape model from which samples can be drawn can be inserted here. We will use the popular linear point distribution models (PDM) as proposed by Cootes and Taylor [1] to model the object shape variations observed in a training set.

Shapes are defined by the coordinates of a set of landmark points which correspond between different shape instances. A principal component analysis on a collection of aligned example shapes yields the so-called modes of shape variation which describe a joint displacement of all landmarks. Each shape can then be approximated by a linear combination of the mean shape and these modes of variation. Usually only a small number of modes is needed to capture most of the variation in the training set.

4.2 Neighbor interaction

As was described in Section 3, the interaction between neighboring shapes can be introduced in the step of perturbation of the degenerate particle set after resampling. This requires $P(S_1|S_2)$, the probability distribution of the expected neighbor of a given shape. In the case where both shapes are modelled with a linear PDM, this is given by the Gaussian conditional density

$$P(S_1|S_2) = \mathcal{N}(\mu, K)$$

with

$$\mu = \Sigma_{12}\Sigma_{22}^{-1}S_2$$
$$K = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

and Σ_{ij} are obtained from the covariance matrix of the combined model

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

as

$$\Sigma_{ij} = \frac{1}{n-1}\sum_n (S_{in} - \bar{S}_i)(S_{jn} - \bar{S}_j)^T.$$

Alternatively, one could leave out the interaction between particles (i.e. perturbed particles are always similar to the particle that produced them) to obtain a segmentation of an unknown number of objects with unknown spatial interrelations.

Both approaches are tested on the rib data.

4.3 Image observation model

In the following, we will use class probability density measurements rather than discrete classes as the image observations as given in Equation 2. The class probability is obtained using a pixel classifier trained to distinguish between foreground and background pixels on the basis of local image descriptors. We have used a k-NN classifier and the outputs of a set of Gaussian derivative filters at multiple scales as image features.

Class probabilities in a pixel x are then defined by

$$P(\omega|x) = \frac{k_\omega}{k},$$

where k_ω among the k nearest neighbors of x belong to class ω.

We still assume that the pixel class due to the individual shapes is given deterministically, that is, each shape S_i is associated with a fixed class template $T_i(x, \omega)$ that defines to which class each pixel belongs. Typically, there will be two classes, one object and one background class. The aim is thus to produce a shape distribution, expressed as a weighted set of shape particles, of which the maximum likelihood classification $M(x, \omega)$ is as similar as possible to the observed (soft) classification of the image $C(x, \omega)$.

4.4 Algorithm

The algorithm for the desired optimization over θ and α looks as follows:

- Sample N shape particles S_i randomly from the prior distribution $p(S)$
- Repeat:
  1. Compute weight α_i for each particle ($\sum_i \alpha_i = 1$):
     Initialize:
     Particles S_i receive a weight α_i according to their overlap with the observed classification C, normalized for size. If several particles vote for the same class in the same pixel, they share the weight between them:

$$\alpha_i = \frac{1}{\sum_{x,\omega} T_i(x,\omega)} \sum_{x,\omega} \frac{C(x,\omega) \times T_i(x,\omega)}{H(x,\omega)}$$

$$H(x,\omega) = \sum_s \frac{T_i(x,\omega)}{\sum_{x,\omega} T_i(x,\omega)}$$

     Optimize α_i:
     In random permutation over particles S_i, with decreasing step size $d\alpha$:
     (a) select particle S_i
     (b) increase α_i by $d\alpha$, decrease $\alpha_{\neg i}$ so that $\sum_s \alpha_i = 1$
     (c) $M(x,\omega,\alpha) = \frac{\sum_s \alpha_i T_i(x,\omega)}{\sum_{s,\omega} \alpha_i T_i(x,\omega)}$

(d) $f = \sum_{x,\omega} M(x, \omega, \alpha) \times C(x, \omega)$
(e) if f increased accept new α

2. Produce a new particle set through weighted sampling with replacement according to α_i
3. Perturb the particles from the new sample set by sampling from $p(S|S_i)$

5 Rib segmentation using shape particle filtering

A set of leave-one-out experiments was performed on 30 standard digitized posterior-anterior chest radiographs of size 256×256, taken from the publicly available JSRT (Japanese Society of Radiological Technology) database [11]. This section describes the specific choices made for rib segmentation.

5.1 Shape models

The proposed algorithm can simultaneously detect and segment an unknown number of similar objects. This could include several different types of objects as well.

For the rib application we have constructed two shape models, one for the left ribs and one for the right ribs. The size and shape of a rib in an X-ray image is strongly correlated with the position in the image. We therefore do not perform a full Procrustes alignment as would usually be preferred in shape model based segmentation. Instead, we translate each image such that the top of the lung fields (minimal y coordinate) and the horizontal center (median x-coordinate) coincide. Since the task of lung field segmentation is much less cumbersome than rib segmentation [8], we will assume that these coordinates are (approximately) known in a new image.

The lung fields and the part of the ribs that is visible in the lung fields have been manually delineated in all images. Ribs are subsequently described by landmarks equidistantly interpolated between the four corner points where a rib intersects the lung field. Ribs that have fewer than 4 corner points are not taken into account in model construction, but they can still be segmented as a variation in position of the ribs is automatically included in the model.

The spatial relations that are modelled are the first neighbor relationships between consecutive ribs in the same lung field and between the ribs that are at the same height in the both lung fields. Thus, in the particle perturbation step, a rib shape from the shape collection can either produce a perturbed version of itself or of its upper, lower, or left/right neighbor.

A linear PDM of the two shapes concatenated in one shape vector is constructed as described in Section 4. The models for a single rib, the combined models and an example of a conditional model as used in neighbor interaction are shown in Figure 1.

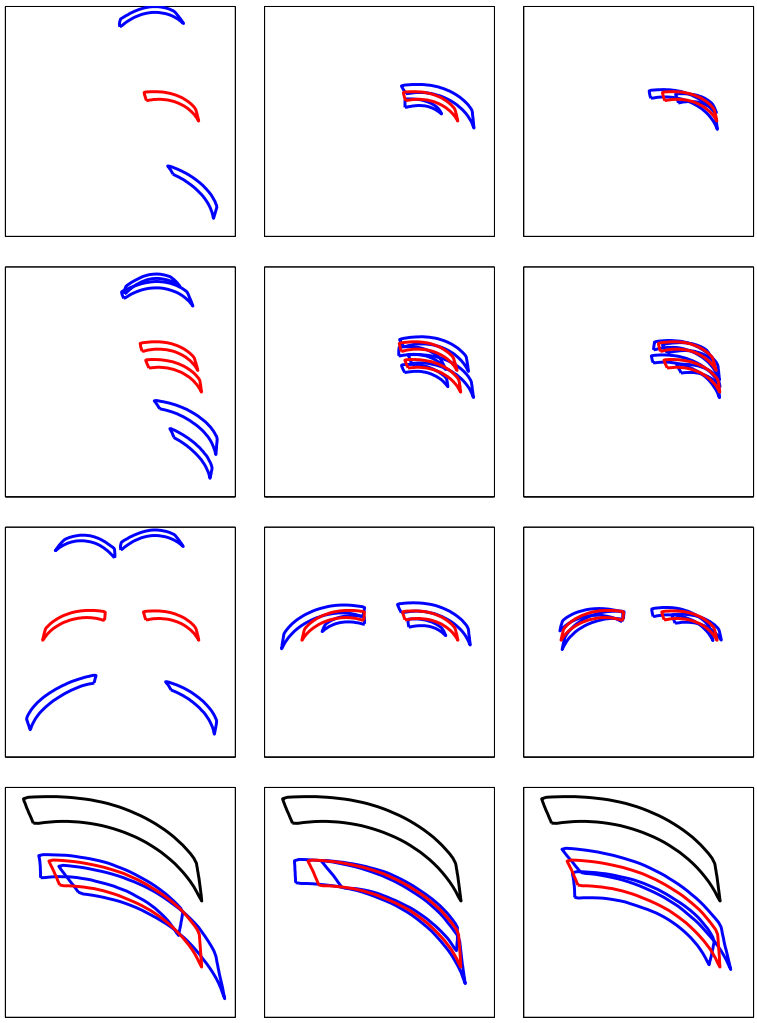

Fig. 1. Examples of the shape-and-pose models constructed for rib segmentation. From left to right, the first three modes of shape variation are visualized with the mean shape in red and the mean shape ± 2 standard deviations in blue, except for row 4 which shows the mean shape ± 4 standard deviations. From top to bottom: 1. Right rib model 2. Right successive ribs model 3. Opposite ribs model 4. Model of the lower rib conditioned on the mean shape of the upper rib. The axes of the plots correspond to the true image size for rows 1 – 3; row 4 is a close-up.

5.2 Settings

We use a set of Gaussian derivative filters at multiple scales as image features and a k-NN classifier for probability estimation. Features include the original image and the derivatives up to the third order computed at a scale of 1, 2, and 4 pixels, resulting in a 21 dimensional feature space. The set of samples is normalized to unit variance for each feature, and k-NN classification is performed with an approximate k-NN classifier [12] with k=25. These settings were selected because they previously yielded good results on lung field classification [7] and have not been adjusted for rib classification.

Class templates as defined for each shape have two classes; inside the rib and outside the rib but within the lung fields. The template is defined by the interior of a rib shape plus a border of 5 background pixels (approximately half the thickness of a rib), so that most of the ribs can be described without overlapping the rib class of one shape template with the background class of another shape's template.

In the experiments presented here, the algorithm was run for 10 iterations without checking for convergence. The number of particles used for filtering is 1000, starting with 500 left ribs and 500 right ribs. The noise added in the particle perturbation step is of standard deviation $\sigma_d = 0.05\,\sigma$, with σ the standard deviation of the prior shape models. The prior for producing itself or one of its three first neighbors is chosen as uniform; each case occurs with a probability of 0.25.

6 Results

An example of segmentations obtained, with and without neighbor interactions, is given in Figure 2. Overall, segmentations using shape set filtering are spatially more consistent than the original pixel classification which includes spurious pixels and shows holes in the ribs. Without neighbor interaction, shape filtering may overlap crossing rib shapes to reconstruct the holes in the original classification. Shape set filtering with neighbor interaction finds the correct consistent rib pattern in most cases.

The error rate of shape set filtering with neighbor interaction is 18.7%, which is not significantly different from the error for the original pixel classification (19.1%, $p = 0.6$). Shape collection filtering without neighbor interaction performs worse (21.7%, $p < 0.0001$).

Figure 3 shows an example of the type of misclassifications by standard pixel classification compared to the proposed method. In general, shape filtering makes fewer gross errors like missing a rib completely or classifying the clavicles as ribs. There are, however, more errors near the rib boundaries which indicates that incorrect shapes have been forced on the segmentation. This may be either caused by an incorrect shape model or by a too strong neighbor interaction. Furthermore, the ribs in the lung tops — which is a problematic area for pixel classification — are difficult to segment also with our method.

The process of evolving the particle set is illustrated in Figure 4.

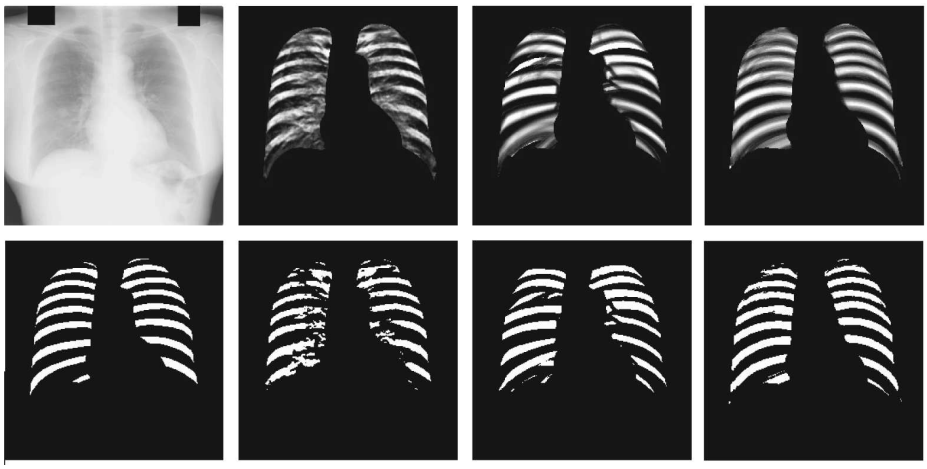

Fig. 2. Examples of segmentations obtained. The top row shows the original X-rays and the different soft classifications; the bottom row shows the hard classifications obtained by thresholding the soft classification at 0.5. From left to right: Ground truth, original pixel classification, shape set filtering without neighbor interaction; shape set filtering with neighbor interaction.

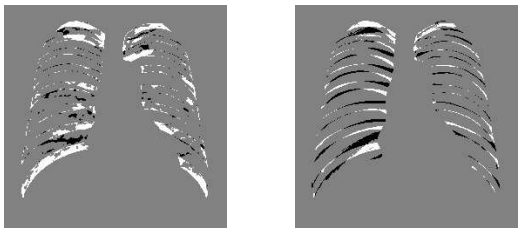

Fig. 3. Typical segmentation errors by pixel classification (left) and shape set filtering with user interaction (right). False positives are in black, false negatives in white, correct classification in gray.

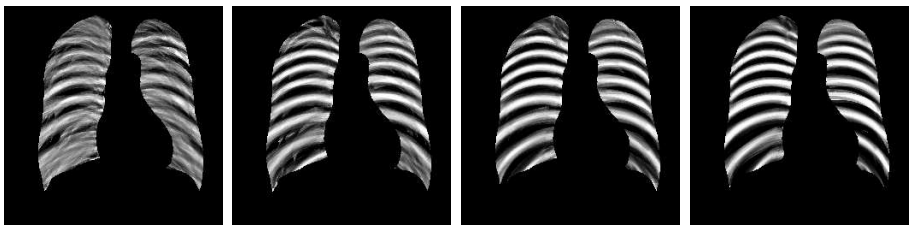

Fig. 4. Evolution of the shape set classification. From left to right: First, third, fifth, and tenth iteration.

7 Discussion

In the current paper we have optimized a collection of shapes on the output of a pixel classifier based on local image descriptors. Such an approach was shown to be successful for single-object segmentation in several medical imaging applications [7]. The incorporation of shape constraints improves the spatial coherence of the pixel classification. However, as the shapes try to adhere to the pixel classification, the results will not be correct if the initial pixel classification results are far from the correct solution. An iterative method in which the method presented here, optimizing a shape collection to match the classification, is alternated with a pixel classification step in which the current shape collection is used as a prior, would likely improve the results.

Further improvements can be expected if more advanced shape models are used. Currently, a large variation in rib shape and position is modelled with a simple linear model, without optimizing point correspondences. This frequently results in 'illegal' shapes being produced. Although this problem is less severe in the case of a large set of shapes, it may result in a blurred and less consistent classification, especially in the less regular top and bottom parts of the lungs.

In addition, in this work the variance in the particle perturbation step was kept constant during iteration. This means that in each iteration new hypotheses of neighboring particles are introduced in the shape collection. If the observed shape did not have a neighbor in that location or the neighbor has a different shape, the weight for these particles will be small, but this still results in a smearing out of the end result and occasionally in an extra rib being detected at the top and the bottom of the lung fields. This could be remedied by adjusting the interaction prior to reflect the fact that ribs in the lung tops are less likely to have an upper neighbor than those in the bottom of the lung fields, or by decreasing the variance of the perturbation density $p(S|S_i)$ over time.

Interaction between neighbors is currently realized by sampling in $p(S|S_i)$ after a new sample set of particles has been selected by sampling proportionally to the image likelihood weights α_i. Thus, the weight of a particle is determined by image forces and interaction between particles is achieved only by successful particles producing hypotheses for their neighbors. A stronger constraint of spatial consistency can be enforced by accepting a particle in the next iteration with a probability proportionally to its consistency with the rest of the current shape collection. We are currently investigating the advantages of various schemes.

Although we have for simplicity assumed that the position of the lung fields in the image is known approximately, in a previous paper we successfully applied shape particle filtering to segmentation of the lung fields [7]. Rather than first segmenting the lungs and subsequently finding ribs *near* the lungs the two tasks could be elegantly combined by filtering ribs and lungs simultaneously where the rib model is conditioned on the lung shapes. This would yield a more constrained shape-and-pose model for the ribs and may lead to better segmentations.

8 Conclusions

We propose a stochastic optimization algorithm which is capable of segmenting an unknown number of similar objects in an image. This method finds spatially more consistent segmentations than pixel classification without shape constraints. Interaction between neighboring shapes enforces consistency in regular patterns of similar shapes and improves upon the results without interaction in segmenting the ribs in chest radiographs.

Acknowledgments

The authors would like to thank B. van Ginneken and M. Loog of the Image Sciences Institute, Utrecht, The Netherlands, for providing the data sets and manual segmentations used in this study.

References

1. T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models – their training and application," *Computer Vision and Image Understanding* **61**(1), pp. 38–59, 1995.
2. T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(6), pp. 681–684, 2001.
3. A. Jain, Y. Zhong, and M. Dubuisson-Jolly, "Deformable template models: A review," *Signal Processing* **71**(2), pp. 109–129, 1998.
4. M. Isard and A. Blake, "Visual tracking by stochastic propagation of conditional density," in *ECCV*, *LNCS* **1064**, pp. 343–356, Springer, 1996.
5. A. Doucet, N. de Freitas, and N. Gordon, eds., *Sequential Monte Carlo methods in practice*, Springer-Verlag, New York, 2001.
6. J. Sullivan, A. Blake, M. Isard, and J. MacCormick, "Object localization by Bayesian correlation," in *ICCV*, pp. 1068–1075, IEEE Computer Society Press, 1999.
7. M. de Bruijne and M. Nielsen, "Shape particle filtering for image segmentation," in *MICCAI*, *LNCS* **3216**, pp. I:186–175, Springer, 2004.
8. B. van Ginneken, B. ter Haar Romeny, , and M. Viergever, "Computer aided diagnosis in chest radiography: A survey," *IEEE Transactions on Medical Imaging* **20**(12), pp. 1228–1241, 2001.
9. M. Loog, *Supervised Dimensionality Reduction and Contextual Pattern Recognition in Medical Image Processing*. PhD thesis, Utrecht University, 2004.
10. J. Ramachandran, M. Pattichis, and P. Soliz, "Pre-classification of chest radiographs for improved active shape model segmentation of ribs," in *Southwest Symposium on Image Analysis and Interpretation*, IEEE Computer Society Press, 2003.
11. J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, , and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating character- istic analysis of radiologists' detection of pulmonary nodules," *American Journal of Roentgenology* **174**, pp. 71–74, 2000.
12. S. Arya, D. Mount, N. Netanyahu, R. Silverman, and A. Wu, "An optimal algorithm for approximate nearest neighbor searching," *Journal of the ACM* (45), pp. 891–923, 1998.