

# Shape Particle Filtering for Image Segmentation

Marleen de Bruijne and Mads Nielsen

IT University of Copenhagen, Denmark

**Abstract.** Deformable template models are valuable tools in medical image segmentation. Current methods elegantly incorporate global shape and appearance, but can not cope with localized appearance variations and rely on an assumption of Gaussian gray value distribution. Furthermore, initialization near the optimal solution is required.

We propose a maximum likelihood shape inference that is based on pixel classification, so that local and non-linear intensity variations are dealt with naturally, while a global shape model ensures a consistent segmentation. Optimization by stochastic sampling removes the need for accurate initialization.

The method is demonstrated on three different medical image segmentation problems: vertebra segmentation in spine radiographs, lung field segmentation in thorax X rays, and delineation of the myocardium of the left ventricle in MRI slices. Accurate results were obtained in all tasks.

## 1 Introduction

Statistical models of global object appearance are widely used for image segmentation [7, 13, 12], and form powerful tools especially in the case of missing or locally ambiguous boundary evidence.

However, entirely global models can be too constrained to adhere to new images adequately. Intensity variations occurring at random locations within an object, such as for example calcification or lesions, can not be captured in a global appearance model and will impair the model fit. To keep model complexity within bounds usually a simple linear model of appearance is applied and thus results are unreliable if the image gray values are not Gaussian distributed. Linear models of object appearance were shown to fail in many medical image segmentation tasks [18, 9, 4, 14]. Another drawback of current deformable model approaches is that they require initialization near the final solution, and thus need manual intervention [9, 16] or automatic object recognition [5, 20].

Suggested solutions for region segmentation of images with a non-linear appearance are based on non-linear filtering or normalization of the images before applying the appearance model [4, 14]. This overcomes some of the problems related to non-Gaussian distributed gray values, but the application to different types of distributions is still rather limited. In edge-based segmentation, non-linear appearance has been modeled as a mixture of Gaussians [5], or by using non-parametric classifiers to discriminate between object and background pixels [18] or between boundary and non-boundary pixels [9]. The classifier-based approaches can cope with arbitrary gray value distributions but can not directly be extended to a region appearance model due to the computational complexity and the amount of data needed.

On the other hand, performance of pixel classification methods was shown to improve by adding global information, for instance in the form of spatially varying priors obtained from digital atlases [19]. These methods rely on the (rigid or elastic) matching of an atlas to the image, and therefore requires the image appearance to be fairly consistent in the entire image.

We propose a shape model inference on the basis of pixel classification. Localized intensity variations are thus dealt with naturally, while the global shape model ensures a consistent segmentation. By applying a non-parametric classifier we will be able to cope with arbitrary gray value distributions. We show that the maximum likelihood solution can be found using iterated likelihood weighing, which can be implemented using a particle filtering scheme such as are often used in object tracking (see for instance [11]). This makes the segmentation result relatively independent of the initialization, guarantees convergence provided that enough samples are used, and allows for straightforward extension to multi-modal shape models or multiple solutions.

## 2 Maximum Likelihood Shape Inference

Several different schemes for efficient optimization of shape models or combined models of shape and appearance have been proposed, usually resulting in a local maximum likelihood or maximum a posteriori (MAP) estimate. In this work, we will estimate the maximum likelihood rather than the MAP solution, since the latter has an inherent bias towards the local mode of the prior.

In the following we will prove that a maximum likelihood estimate of the shape may be obtained through an iterated likelihood weighing that may simply be implemented by a particle filtering scheme. We will show global convergence to the global maximum.

**Theorem 1 (Maximum likelihood convergence).** *The maximum of  $P(I|S)^t P(S)$  converges to the maximum of  $P(I|S)$  for any  $P(S)|\forall S, P(S) > 0$ , and  $S \in \mathcal{B}$ , where  $\mathcal{B}$  is a separable Banach space endowed with a weak (or strong) topology.*

*Proof.* Let us define

$$f_t(S) = (P(I|S)^t P(S))^{1/t}$$

Since  $f_t(S)$  is just a monotonic transformation of  $P(I|S)^t P(S)$  it has the same properties of extrema, and we may examine  $f_t$  instead. Since  $f_t$  converges to  $P(I|S)$  uniformly (remember  $P(S) > 0$ ), it also  $\Gamma$ -converges to  $P(I|S)$  ([3] page 35, Theorem 2.1.8). Since  $\Gamma$ -convergence is obtained,

$$\lim_{t \rightarrow \infty} \arg \max_S f_t(S) = \arg \max_S P(I|S)$$

([3] page 34, Theorem 2.1.7)

Hence we have proved that the maximum likelihood shape may be obtained as the maximum of  $P(I|S)^t P(S)$  as  $t$  approaches infinity. Let us define

$$L_t(S) = \frac{P(I|S)^t P(S)}{Z_t}$$

where  $Z_t = \int P(I|S)^t P(S) dS$ . As  $t$  tends to infinity  $L_t(S) \rightarrow \sum_i a_i \delta(S_i)$ , where  $\delta$  is the Dirac delta functional,  $a_i$  are positive constants summing to 1, and  $S_i$  is the shapes taking the global maximum value  $P(I|S)$ . Especially in the case where  $P(I|S)$  has a unique maximum,  $Z_t$  tends to a single  $\delta(S_0)$ , where  $S_0$  is the maximum likelihood estimate.

Hence we have shown that samples of  $L_t(S)$  with probability 1 tends to the maximum likelihood estimate of  $S$ .

Notice that the above arguments could have been made simpler if  $S$  belongs to a normed space with a strong topology (like  $R^N$ ). However, the above results are valid for shapes defined as points in Kendall's shape space, closed curves, surfaces in 3D, functions of bounded variation on the real plane, etc. That is, any shape representation living in a Banach space (a complete normed linear space) equipped with a weak or strong topology.

## 2.1 Shape Particle Filtering

As shown, iterated likelihood weighing of a shape distribution leads to the maximum likelihood shape solution. We will implement this using a particle filtering scheme [10]. It must be noted, that the argumentation in the previous section holds in the continuous domain; in a discretized shape space and within a finite evolution time the maximum likelihood solution may not be reached. However, using a sufficiently dense sampling of the shape distribution and a sufficiently slow evolution, the algorithm will converge.

We will use a shape model derived from hand annotated examples for  $P(S)$ , and estimate the likelihood  $P(I|S)$  using pixel classification. Within this framework, any kind of shape model, any set of local image descriptors and any classifier can be used.

**Shape Model.** The object shape and shape variation are described using a point distribution model (PDM) [8]. Shapes are defined by the coordinates of a set of landmark points which correspond between different shape instances. Each shape can be approximated by a linear combination of the mean shape and several modes of shape variation which describe a joint displacement of all landmarks. The modes of variation are given by the principal components of a collection of aligned example shapes. Usually only a small number of components is needed to capture most of the variation in the training set.

**Image term.** Each shape has associated with it a labeling that divides the image pixels into two or more classes, for example inside and outside an object. A pixel classifier is trained to distinguish between pixels of different classes on the basis of local image descriptors. We have chosen a general scheme in which pixels are described by the outputs of a set of Gaussian derivative filters at multiple scales, and a k-NN classifier is used for probability estimation.

We use a moderated k-NN classifier by which the posterior probability of a pixel with feature vector  $\mathbf{x}$  belonging to class  $\omega$  is given by

$$P(\omega|\mathbf{x}) = \frac{k_\omega + 1}{k + m},$$

where  $k_\omega$  among the  $k$  nearest neighbors belong to class  $\omega$ , and  $m$  is the number of classes [1]. The moderation with respect to a standard  $k$ -NN classifier with  $P(\omega|\mathbf{x}) = k_\omega/k$  ensures that probabilities are always nonzero for finite  $k$ , thus avoiding a ‘veto-effect’ if probabilities of separate pixels are multiplied.

Pixel intensities are assumed to be conditionally independent on the class label, and the likelihood of a shape template is thus given by the product of separate pixel likelihoods. Particle weights are then defined as

$$w_i = \exp \left[ \frac{c}{n} \sum_{j=1}^n \log P(\mathbf{x}_j | \omega_{s_i}) \right], \quad (1)$$

where  $c$  is a constant which controls the randomness of the sampling process,  $n$  is the number of pixels in the template, and  $P(\mathbf{x}_j | \omega_{s_i})$  is the likelihood term for the observed pixel feature vector  $\mathbf{x}_j$  given the implied label  $\omega_{s_i}$ .

**Particle Filtering.** A random set of  $N$  shape hypotheses — ‘particles’ —  $s_i$  is sampled from the prior shape-and-pose model. Each hypothesis has an associated image labeling, which is compared to the label probability map as obtained from the initial pixel classification. Particles are weighed by their likelihood term and a new set of  $N$  hypotheses is generated from the current set by random sampling proportionally to these weights. In this way, particles representing unlikely shapes vanish while successful particles multiply. After a small random perturbation of duplicate particles, the process of importance resampling is repeated. The initial sparse sampling eventually evolves into a  $\delta$ -peak at the maximum likelihood solution.

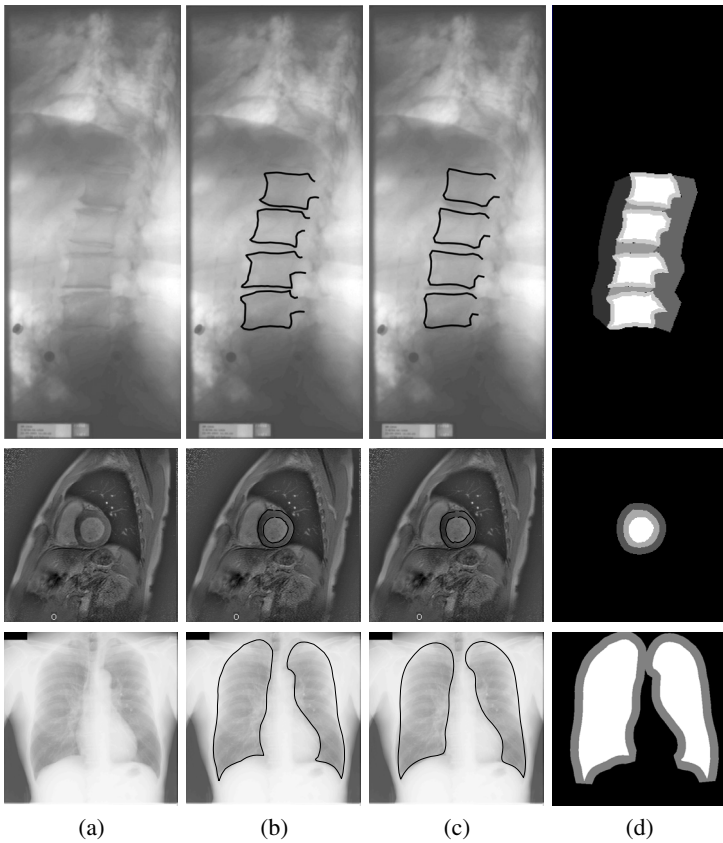
In practice, the solution may be approximated by the strongest local mode of the particle distribution before the process has converged. The local modes can be efficiently obtained using the mean shift algorithm [6].

### 3 Experiments

Cross-validation experiments are performed on spine and thorax radiographs and on short-axis cardiac MRI slices. Examples of each of these image types, with the desired segmentation, are given in Fig. 1(a-b).

**The spine data set** contains 91 lateral spine radiographs of both healthy and fractured vertebrae, in which the vertebrae L1 through L4 are delineated manually. Image sizes are ca.  $210 \times 505$  pixels. Shape models are constructed by equidistant sampling along the contours of each of the vertebrae. A set of cross-validation experiments is performed in which each time the method is trained on 84 images and tested on the remaining 7.

**The lung data set** contains 30 digitized posterior-anterior chest radiographs of size  $512 \times 512$ , taken from the publicly available JSRT (Japanese Society of Radiological Technology) database [15]. Both lung fields have been delineated manually by two observers. Shape models are constructed by equidistant sampling of each lung, starting from the outer and lower corner. The left and right lungs are modeled independently. Leave-one-out experiments are performed.



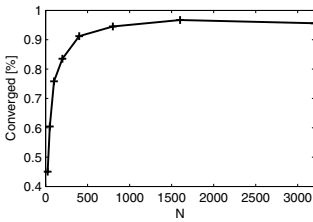
**Fig. 1.** Example images from each data set. Top row: spine X ray; middle row: cardiac MRI; bottom row: thorax X ray. (a) Normalized image; (b) Manual segmentation; (c) Automatic segmentation using shape particle filtering; (d) Label template used.

**The cardiac data set** was made publicly available by M.B. Stegmann [17] and consists of 14 short-axis, end-diastolic cardiac MRI slices of size  $256 \times 256$  with manually placed landmarks on the epicardial and endocardial contours. Leave-one-out experiments are performed.

All images have been normalized to zero mean and unit variance prior to processing, using a global normalization for the X rays, and for the MR images a Gaussian local normalization according to  $\bar{L} = (L - L_\sigma) / \sqrt{(L^2)_\sigma - (L_\sigma)^2}$ , where  $L_\sigma$  is the image  $L$  convolved with a Gaussian kernel of width  $\sigma$  ( $\sigma=16$  pixels).

### 3.1 Settings

The parameter settings used for pixel classification were the same for all experiments: Features include the original image and the derivatives up to the third order computed



(a)

	$N$	mean	stdev	max
Spine	1600	1.4	0.4	3.1
Cardiac	1000	1.1	0.3	1.9
Left lung	2000	4.4	2.0	12
Right lung	2000	3.8	1.5	9.8
Left lung, 2nd observer		3.6	3.8	23
Right lung 2nd observer		3.0	1.7	8.4

(b)

**Fig. 2.** (a) Percentage of spine segmentations converged with an average error below 2 mm, as a function of the number of particles in the distribution  $N$ . (b) Average point-to-contour segmentation errors in pixels.

at a scale of 1, 2, 4, and 8 pixels, resulting in a 41 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 [2] with  $k=25$ .

For each task, class templates have been defined from the manually drawn contours. For lung and heart segmentation, this is simply the interior of the contour, and a border on the outside. In the spine data one can, on the basis of the shape alone, make a distinction between parts of the image with a different appearance, without requiring additional annotation. We have defined a template with five separate image regions: anterior background, posterior background, inter vertebral space, vertebra boundary and vertebra interior (see Fig. 1(d)).

Apart from  $N$ , the number of shape particles used in the process, all particle filtering parameters are equal in all experiments. The noise added to duplicates in the particle filtering process is of standard deviation  $\sigma_d = 0.05 \sigma$ , with  $\sigma$  the standard deviation of the prior shape model. Iteration stops if the change in the maximum local mode of the shape distribution, computed using mean shift filtering with a kernel of size  $0.05 \sigma$ , becomes negligibly small. The constant  $c$ , determining the speed of convergence according to Eq. 1, is set to 50.

### 3.2 Results

Examples of segmentations obtained are given in Fig. 1(c). Segmentation accuracy is evaluated by the mean point-to-contour distance with respect to the manual outlines Fig. 2(a) shows performance of vertebra segmentation for particle distributions of varying size. Results for a fixed number of particles and for all three data sets are listed in Fig. 2(b). In the case of vertebra segmentation, where the model describes only a part of the entire spine, a shape instance shifted one or more vertebra heights upwards or downwards with respect to the manual segmentation can also represent accurate vertebra segmentations. We allow the model to shift one vertebra and report the errors of the three overlapping vertebrae.

## 4 Discussion and Conclusion

The obtained accuracy of vertebra segmentation is close to the maximum accuracy that can be expected with the given shape model, and results are likely to improve if higher resolution images, a higher dimensional shape model, and more training images are used. Nevertheless, the results are competitive with results described in the literature. Zamora et al. reported success rates of 50% with average errors below 6.4 mm in active shape model (ASM) segmentation of lumbar vertebrae in spine radiographs [20]. Smyth et al. performed ASM segmentation of vertebrae in dual energy X ray absorptiometry (DXA) images [16] and obtained success rates of 94–98%, with errors in the order of 1 mm for healthy vertebra and success rates of 85–98% with errors in the order of 2 mm for fractured vertebrae. Scott et al. reported successful convergence of a modified active appearance model (AAM) in 92% of DXA scans of healthy spines with an average error of ca. 1.5 mm [14].

Segmentation of cardiac MRI is a task where linear AAMs have been shown to perform well. Stegmann et al. [17] reported point-to-contour errors of basic AAM of 1.18 (max 2.43), on the same data set. If a border of background appearance was added to the model, which usually improves segmentation performance, the mean error increased to 1.73. The mean errors obtained using shape particle filtering, with a border around the template, are comparable to those of the basic AAM ( $1.1 \pm 0.3$ , max 1.9). Without a border we obtained similar errors ( $1.2 \pm 0.3$ , max 2.1), which indicates that our method is less sensitive than AAM to the large variation in intensity of tissue surrounding the left ventricle, even though these different tissues end up in the same class of ‘background tissue’.

The errors for lung field segmentation are a little larger than those of the other two tasks, but inter observer variation is also large in these images. It must be noted that the placement of landmarks, equidistantly along the contour with only one specific corresponding point, is far from optimal in this case and an optimization of landmark positioning will likely improve segmentation results.

The use of a large number of hypotheses makes segmentation by shape particle filtering robust to local maxima and independent of initialization. An additional advantage of particle filters, not employed in this work, is their ability to represent multiple solutions simultaneously. This could for instance be used to segment the entire spine with only a partial spine model. Furthermore, possible multimodal shape distributions would be dealt with naturally.

To conclude, we propose a robust and general method for the segmentation of images with localized or non-linear gray value variations.

**Acknowledgements.** The authors would like to thank L. Tánko and C. Christiansen of the Center for Clinical and Basic Research (CCBR A/S), Denmark, J.C. Nilsson and B.A. Grønning of the Danish Research Center for Magnetic Resonance (DCRMR), M.B. Stegmann of the Technical University of Denmark and 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. F.M. Alkoot and J. Kittler. Moderating k-NN classifiers. *Pattern Analysis & Applications*, 5(3):326–332, 2002.
2. S. Arya, D.M. Mount, N.S. Netanyahu, R. Silverman, and A.Y. Wu. An optimal algorithm for approximate nearest neighbor searching. *Journal of the ACM*, (45):891–923, 1998.
3. Gilles Aubert and Pierre Kornprobst. *Mathematical problems in image processing*, volume 147 of *Applied Mathematical Sciences*. Springer Verlag, 2002.
4. H.G. Bosch, S.C. Mitchell, B.P.F. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J.H.C. Reiber. Active appearance-motion models for endocardial contour detection in time sequences of echocardiograms. In *Med Imaging: Image Process*, volume 4322 of *Proc of SPIE*. SPIE Press, 2001.
5. M. Brejl and M. Sonka. Object localization and border detection criteria design in edge-based image segmentation: automated learning from examples. *IEEE Trans Med Imaging*, 19(10):973–985, 2000.
6. D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE TPAMI*, 24(5):603–619, 2002.
7. T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. *IEEE TPAMI*, 23(6):681–684, 2001.
8. T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Active shape models – their training and application. *Comput Vis Image Underst*, 61(1):38–59, 1995.
9. M. de Bruijne, B. van Ginneken, M.A. Viergever, and W.J. Niessen. Adapting active shape models for 3D segmentation of tubular structures in medical images. In *IPMI*, volume 2732 of *LNCS*, pages 136–147. Springer, 2003.
10. A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo methods in practice*. Springer-Verlag, New York, 2001.
11. M. Isard and A. Blake. CONDENSATION – conditional density propagation for visual tracking. *Int J Comput Vis*, 29(1):5–28, 1998.
12. A.K. Jain, Y. Zhong, and M.P. Dubuisson-Jolly. Deformable template models: A review. *Signal Processing*, 71(2):109–129, 1998.
13. S. Sclaroff and J. Isidoro. Active blobs: region-based, deformable appearance models. *Comput Vis Image Underst*, 89(2-3):197–225, 2003.
14. I.M. Scott, T.F. Cootes, and C.J. Taylor. Improving appearance model matching using local image structure. In *IPMI*, volume 2732 of *LNCS*, pages 258–269. Springer, 2003.
15. 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 characteristic analysis of radiologists’ detection of pulmonary nodules. *Am J Roentgenol*, 174:71–74, 2000.
16. P.P. Smyth, C.J. Taylor, and J.E. Adams. Vertebral shape: Automatic measurement with active shape models. *Radiology*, 211(2):571–578, 1999.
17. M.B. Stegmann, R. Fisker, and B.K. Ersbøll. Extending and applying active appearance models for automated, high precision segmentation in different image modalities. In *Proceedings of the 12th Scandinavian Conference on Image Analysis*, 2001.
18. B. van Ginneken, A.F. Frangi, J.J. Staal, B.M. ter Haar Romeny, and M.A. Viergever. Active shape model segmentation with optimal features. *IEEE Trans Med Imaging*, 21(8):924–933, 2002.
19. S.K. Warfield, M. Kaus, F.A. Jolesz, and R. Kikinis. Adaptive, template moderated, spatially varying statistical classification. *Med Image Anal*, 4:43–55, 2000.
20. G. Zamora, H. Sari-Sarrafa, and R. Long. Hierarchical segmentation of vertebrae from x-ray images. In *Med Imaging: Image Process*, volume 5032 of *Proc of SPIE*, pages 631–642. SPIE Press, 2003.