Image Segmentation by Shape Particle Filtering

Marleen de Bruijne and Mads Nielsen IT University of Copenhagen, Denmark

Abstract

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

We propose a shape inference method 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 vertebra segmentation in spine radiographs. Segmentation errors are below 2 mm in 88 out of 91 cases, with an average error of 1.4 mm.

1. Introduction

Deformable templates of global object appearance are widely used for image segmentation [8, 6, 13, 11]. These techniques can produce correct results even 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 deformable templates usually apply a simple linear model of appearance and thus produce unreliable results if the image gray values are not Gaussian distributed. Linear models of object appearance were shown to fail in many medical image segmentation tasks [16, 9, 3, 14].

Another drawback of current deformable model approaches it that they require initialization near the final solution, and thus need manual intervention [9, 15] or automatic object recognition [4, 19].

Several authors have proposed hierarchical frameworks in which a global shape and/or appearance model is used for object localization, whereupon additional local deformations are modeled using snakes [4, 19], Markov processes [12], or small versions of the global appearance model [7].

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 [3, 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 [4], or by using non-parametric classifiers to discriminate between object and background pixels [16] 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, the result of previously purely local segmentation using pixel classification was shown to improve by adding global information in the form of spatially varying priors obtained from digital atlases [17, 18]. 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.

The contribution of this work is twofold. First, a shape model is optimized on the output of a pixel classification step. Local intensity variations are dealt with naturally, while the global shape model ensures a consistent segmentation. By applying a non-parametric classifier the algorithm can cope with arbitrary gray value distributions. Second, shape inference is performed using a stochastic sampling technique. 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. Shape Particle Filtering

The proposed segmentation scheme requires a global shape and a local appearance model, which are both derived



Figure 1. Example of optimization by particle filtering, showing the first three components of the shape-and-pose distribution. (a) Random initialization with 1600 particles; (b-d) Distribution after (b) one, (c) three, and (d) five iterations. The shape likelihood clearly follows a multimodal distribution, where distinct maxima correspond to instances of the four-vertebra model shifted along the spine.

from hand annotated example images. Each shape is associated with one unique labeling of the image 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.

In a new image, first a probability map for each label is computed using the classifier. The optimal shape, given the shape prior and the initial classification, is then obtained using a random sampling technique similar to the *particle filtering* that is often used in object tracking [10].

Pixel Classification Within this segmentation framework, any set of local image descriptors and any classifier can be used for the appearance model, dependent on the demands on segmentation speed and accuracy and on the type of images. 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 to estimate the probability non-parametrically.

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

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

where k_{ω} among the k nearest neighbors belong to class ω , and m is the number of classes [1].

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, in the ideal case, denote the same anatomical points in different 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.

Particle Filtering A random set of N shape hypotheses — 'particles' — s_i is sampled from the prior shape-andpose 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 thus evolves into a distribution with high density around the most likely shapes. This process is illustrated in Figure 1.

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

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

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

The optimal fit is given by the strongest local mode of the particle distribution, obtained with the mean shift algorithm [5]. Iteration stops if the change in shape between iterations becomes negligible.

3. Experiments

Experiments are performed on 91 lateral spine radiographs taken from an osteoporosis screening program. The dataset contains both healthy and fractured vertebrae. The vertebrae L1 through L4 are delineated manually by a medical expert. The original radiographs have been scanned,



Figure 2. (a) Original image; (b) Labeling implied by the manually drawn outline of L1-L4 vertebrae; (c-g) Classification results for the different labels: (c) anterior background; (d) posterior background; (e) inter vertebral space; (f) vertebra boundary; (g) vertebra interior; (h) Resulting soft classification into vertebra and background; (i) Final segmentation result.

normalized to zero mean and unit variance, and downsampled to a resolution of 0.85 mm per pixel. A set of crossvalidation experiments is performed in which each time the classifier is trained on 81 images and tested on 10.

Setup for vertebra segmentation One shape model is constructed for all four vertebrae together, using an equal number of landmarks (25) placed equidistantly along the outlines. Shapes are aligned such that their centers of gravity coincide and the first four modes of remaining variation, which describe 89% of the total variation in the training set, are selected.

We define a template with five image regions that look roughly the same in all images: anterior background, posterior background, inter vertebral space, vertebra boundary and vertebra (see Figure 2). The classifier is trained on one out of ten pixels that lie within the template, resulting in a total of around 130.000 training samples. Features include the original images and the derivatives up to the third order computed 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.

The noise added to duplicates in the particle filtering process is of standard deviation $\sigma_d = 0.05 \sigma$, with σ the standard deviation of the prior shape model. The size of the kernel for local mode finding with the mean shift algorithm is 0.05σ as well.

Results An example of results obtained is given in Figure 2. From Figure 2.h it is clear that pixel classifica-

tion alone does not provide enough information for measurements of vertebra shape; part of the spine is identified correctly but many false positives and false negatives are present.

The segmentations obtained are evaluated by the mean distance from the landmarks to the expert-drawn contour. Since the shape model describes only part of the spine, a model shifted one or more vertebra lengths upwards or downwards can also represent accurate vertebra segmentations. A final application should use a full spine model, allow for multiple solutions, or include e.g. part of the pelvis bone in the model. In this work, we allow the model to shift one vertebra and in that case compute the errors of the three overlapping vertebrae only.

Figure 3 shows the segmentation results for particle distributions of varying size. At N = 1600, errors were below 2 mm in 97% of cases, with an average error of 1.4 mm. The mean error of the optimal fit of the shape model to the manual contours was 1.1 mm.

4. Discussion and Conclusion

The obtained segmentation accuracy 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, feature selection, more samples in the k-NN classifier, and more training images are used.

Nevertheless, the results are competitive with results described in the literature. Zamora et al. reported average errors below 6.4 mm in 50% of cases in active shape model (ASM) based segmentation of lumbar vertebra in spine radiographs [19]. Smyth et al. performed ASM segmentation



Figure 3. Segmentation results as a function of N, the number of particles in the distribution. Left: Percentage of results converged with an average error below 4 mm (dotted) or 2 mm (solid line). Right: Average error of accepted results.

of lumbar vertebrae in dual energy X ray absorptiometry (DXA) images [15] 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].

The proposed combination of global shape and local appearance models will perform best if globally consistent appearance variation, such as differences in tissue type, are indeed described by the shape model and get a different class label. If this is not the case one could manually define different classes, as was done with the spine model. Alternatively, a subdivision can be obtained using unsupervised clustering on the training set.

The use of a large number of hypotheses makes segmentation by shape particle filtering relatively 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 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, supervised method for the segmentation of images with locally varying or non-linear gray value distributions.

References

[1] F. Alkoot and J. Kittler. Moderating k-NN classifiers. *Pattern Analysis & Applications*, 5(3):326–332, 2002.

- [2] S. Arya, D. Mount, N. Netanyahu, R. Silverman, and A. Wu. An optimal algorithm for approximate nearest neighbor searching. *Journal of the ACM*, (45):891–923, 1998.
- [3] H. Bosch, S. Mitchell, B. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J. 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.
- [4] 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.
- [5] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE TPAMI*, 24(5):603–619, 2002.
- [6] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *IEEE TPAMI*, 23(6):681–684, 2001.
- [7] T. Cootes and C. Taylor. Combining elastic and statistical models of appearance variation. In *Proc. ECCV'00, Part I*, volume 1842 of *LNCS*, pages 149–163. Springer, 2000.
- [8] T. Cootes, C. Taylor, D. 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. Viergever, and W. 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, 2001.
- [11] A. Jain, Y. Zhong, and M. Dubuisson-Jolly. Deformable template models: A review. *Signal Processing*, 71(2):109– 129, 1998.
- [12] C. Kervrann and F. Heitz. A hierarchical statistical framework for the segmentation of deformable objects in image sequences. In *Proc. CVPR'94*, pages 724–728. Computer Society Press, 1994.
- [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. Scott, T. Cootes, and C. Taylor. Improving appearance model matching using local image structure. In *IPMI*, volume 2732 of *LNCS*, pages 258–269. Springer, 2003.
- [15] P. Smyth, C. Taylor, and J. Adams. Vertebral shape: Automatic measurement with active shape models. *Radiology*, 211(2):571–578, 1999.
- [16] B. van Ginneken, A. Frangi, J. Staal, B. ter Haar Romeny, and M. Viergever. Active shape model segmentation with optimal features. *IEEE Trans Med Imaging*, 21(8):924–933, 2002.
- [17] K. van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automatic segmentation of brain tissues and MR bias field correction using a digital brain atlas. In *MICCAI*, volume 1496 of *LNCS*. Springer, 1998.
- [18] S. Warfield, M. Kaus, F. Jolesz, and R. Kikinis. Adaptive, template moderated, spatially varying statistical classification. *Med Image Anal*, 4:43–55, 2000.
- [19] 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.