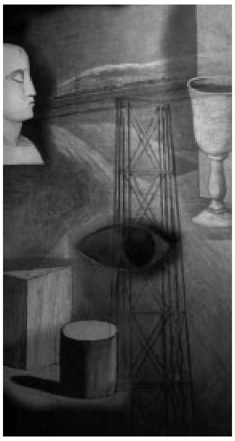
THE JOURNAL OF VISUALIZATION AND COMPUTER ANIMATION
J. Visual. Comput. Animat. 2002; **13:** 1–19
Published online in Wiley InterScience (www.interscience.wiley.com). **(DOI:** 10.1002 / vis.276)

Scale-invariant segmentation of dynamic contrast-enhanced perfusion MR images with inherent scale selection



By J. P. Janssen, M. Egmont-Petersen, E. A. Hendriks, M. J. T. Reinders, R. J. van der Geest, P. C. W. Hogendoorn and J. H. C. Reiber*

Selection of the best set of scales is problematic when developing signal-driven approaches for pixel-based image segmentation. Often, different possibly conflicting criteria need to be fulfilled in order to obtain the best trade-off between uncertainty (variance) and location accuracy. The optimal set of scales depends on several factors: the noise level present in the image material, the prior distribution of the different types of segments, the class-conditional distributions associated with each type of segment as well as the actual size of the (connected) segments. We analyse, theoretically and through experiments, the possibility of using the overall and class-conditional error rates as criteria for selecting the optimal sampling of the linear and morphological scale spaces. It is shown that the overall error rate is optimized by taking the prior class distribution in the image material into account. However, a uniform (ignorant) prior distribution ensures constant class-conditional error rates. Consequently, we advocate for a uniform prior class distribution when an uncommitted, scale-invariant segmentation approach is desired. Experiments with a neural net classifier developed for segmentation of dynamic magnetic resonance (MR) images, acquired with a paramagnetic tracer, support the theoretical results. Furthermore, the experiments show that the addition of spatial features to the classifier, extracted from the linear or morphological scale spaces, improves the segmentation result compared to a signal-driven approach based solely on the dynamic MR signal. The segmentation results obtained from the two types of features are compared using two novel quality measures that characterize spatial properties of labelled images.
Copyright © 2002 John Wiley & Sons, Ltd.

Received: 10 December 2001; Revised: 11 February 2002

KEY WORDS: scale selection; linear scale space; dynamic contrast-enhanced MR imaging; non-linear morphological filtering; morphological scale space; scale-invariant segmentation

Introduction

Segmentation of an image can in many situations be considered a pattern recognition problem where each pixel (or voxel) is to be assigned a specific label by a classifier. Techniques from statistical pattern recognition have in the past been used in many different ways to segment images. Two frequently occurring segmentation tasks are object recognition (e.g., coin recognition[1]) and texture segregation (e.g., feature detection in cartographic images[2]). Different types of approaches have been developed for solving such segmentation problems. An important distinction should be made between approaches that are *signal-driven*,[3] *feature-driven*[4,5] and *modular* approaches;[6] the latter often combine signal- and feature-driven methods. Which of the three types of approach best solves a particular segmentation task depends on the prior knowledge of the problem at hand and of the degrees of freedom that are present in the underlying image material. In many practical segmentation problems,

*Correspondence to: M. Egmont-Petersen, Department of Information and Computing Sciences, Utrecht University, PO Box 80.089, 3508 TB Utrecht, The Netherlands.
E-mail: Michael@cs.uu.nl

concomitant variations in position, orientation and size impede a broad application of the developed approach. In (2D) perspective images, additional degrees of freedom such as slant and tilt often need to be taken into account. In the following, we restrict the scope to segmentation that is invariant with respect to the three affine image transformations: translation, rotation and scaling.

Generally, both signal- and feature-driven approaches cope with variations in position by convolution. A window is slid across the image and its central pixel/voxel is assigned the most likely class label based on, for example, the contents of the window[7] or on a derived feature vector.[8] In the case of non-isotropic patterns, rotation invariance needs to be incorporated into the segmentation algorithm too. A frequently applied technique for incorporating rotation invariance into signal-driven segmentation approaches is preprocessing with the Karhunen–Loève transform (principal component analysis).[1,9,10] For feature-driven segmentation approaches, application of rotation-invariant features—e.g., the moments of Hu,[11] Zernike moments[12] or Fourier descriptors[13]—automatically ensures that the segmentation approach gives the same result regardless of how the image is oriented.

Scale-invariant segmentation is, in general, more difficult to achieve, partly because of the discrete nature of digital images. Appropriate rescaling of a sampled signal requires an interpolation scheme.[14] However, Nyquist's criterion imposes a natural limit to the resolution to which the image can be scaled. Instead of rescaling the image, scale-invariant segmentation can be obtained by, for example, including image patches at different scales in the training set.[15] Another approach entails transforming the image by an invariant mapping such as the Fourier–Mellin transform.[16,17] Scale-invariant segmentation can also be obtained by training a classifier based on features that eliminate changes in scale. Such scale-invariant features include wavelets,[5] features from the linear scale space[18] and different statistical moments.[19,20] Statistical moments have the disadvantage that they are sensitive to noise and distortions. For signal-driven segmentation algorithms that are based on a wavelet decomposition or a stack of scaled images, the classifier needs to learn variations in scale explicitly, which means that changes in scale are regarded as intra-class variation. When a set of features computed at a number of different scales is provided as input to a classifier, not all scales will contribute equally to making the best distinction between the different segments. In previous articles, we suggested to use a feature selection mechanism for identifying the best sampling scheme of the scale space.[9,21] In this article, we will perform a theoretical analysis of the problem of scale selection for signal-driven segmentation algorithms. The derived results are verified by a set of experiments with a sequence of dynamic magnetic resonance (MR) images.

In the following, we will first reformulate segmentation as a classification problem. We then establish a mathematical framework for Bayesian inference in which it is shown how the minimal error rate Bayesian classifier can be used to perform scale selection. Within this framework, we show that selection of the optimal set of scales requires choosing an appropriate trade-off between bias and variance. Moreover, it will be shown that an uncommitted, invariant segmentation algorithm needs to be trained with a uniform prior class distribution. An uncommitted algorithm is desired when there is no information available regarding the prior distribution of the pixels belonging to the different types of segments.

We investigate the theoretical results in a set of experiments where a pattern classifier (neural network) is developed for scale-invariant segmentation of dynamic perfusion MR images. Some theoretical results regarding scale invariance are tested on synthetic images composed of samples from real MR images.

Background

Define an image as a high-dimensional manifold $I(\boldsymbol{x})$, with $\boldsymbol{x}=(x_1, \ldots, x_d)$, $x_i \in X_i$. In the sequel, we first define minimal error rate classification. It is subsequently shown how a minimal error rate classifier can segment an image.

Minimal Error Rate Classification

Let $\boldsymbol{z}$ denote an n-dimensional vector consisting of continuous features. Bayes' classification rule is defined as[22]

$$P(\omega_j|\boldsymbol{z}) = \frac{P(\omega_j)p(\boldsymbol{z}|\omega_j)}{\sum_i P(\omega_i)p(\boldsymbol{z}|\omega_i)} \tag{1}$$

with $P(\omega_j)$, $j=1, \ldots, c$, and $P(\omega_j\,|\,\boldsymbol{z})$ the prior and posterior probabilities that the vector $\boldsymbol{z}$ belongs to class j, respectively, and $p(\boldsymbol{z}\,|\,\omega_j)$ the class-conditional probability density function of class j. In case all misclassifications are considered inducing the same loss, the vector z should be assigned the class label with the maximal posterior probability, $o_j = P(\omega_j \mid z)$:

$$\text{class}(\boldsymbol{o}) = \begin{cases} j: & \forall i \neq j,\ o_i < o_j \\ \min\{j \mid j = \arg\max_i (o_i)\}: & else \end{cases} \tag{2}$$

This classification function is called the winner-takes-all rule and results in a partitioning of the feature space into disjoint regions R_j:[23]

$$R_j = \left\{ z \in \Re^d \middle| P(\omega_j)p(z|\omega_j) > P(\omega_i)p(z|\omega_i),\ \forall i \neq j \right\} \tag{3}$$

The Bayesian classifier that segments the feature space according to the partitioning $R_1, \ldots, R_c$, results in the minimal error rate ε^*:[22]

$$\varepsilon^* = 1 - \sum_j P(\omega_j) P(z \in R_j | \omega_j) \tag{4}$$

with $P(z \in R_j | \omega_j) = \int_{R_j} P(z|\omega_j)\mathrm{d}z$. Define also the class-conditional error rate ε_{ω_j} by

$$\varepsilon_{\omega_j} = 1 - P(z \in R_j | \omega_j) \tag{5}$$

with the conditional probability $P(z \in R_j \mid \omega_j) = P(z \in R_j, \omega_j)/P(\omega_j)$.

In practice, the posterior probabilities $P(\omega_j \mid z)$ are estimated by a classifier that approximates the optimal mapping N: $\Re^n \rightarrow [0,1]^c$, with c the number of classes that need to be discerned.

Signal-Driven Segmentation

Signal-driven segmentation of the image $I(\boldsymbol{x})$ entails a partitioning of the image elements (pixels or voxels) into clusters that correspond to the desired segmentation result. Segmentation can be seen as a classification task, which has as its purpose the assignment of a label to each image element. Define the segmented (labelled) image by an implicit convolution:

$$S(\boldsymbol{x}) = \text{class}(N(I(\boldsymbol{x}))),\ \boldsymbol{x} \in (X_1 \times X_2 \ldots X_d) \tag{6}$$

with $\boldsymbol{o} = N(x)$ denoting the classifier. A connected cluster of image elements in $S(x)$ that are assigned the same label is considered one segment.

Segmentation Using Scale Space Features

In the sequel, we give a brief introduction to the linear scale space and show how a Taylor expansion can be used to capture the local geometric structure of an image.

The Linear Scale Space

A widely applied framework for image analysis that takes explicitly the scale of image features into account, is the linear scale space.[24–27] In the linear scale space, a stack of images is formed as a function of an increasing inner scale t. The two-dimensional linear scale space is based on the *linear diffusion equation*:[25]

$$\frac{\partial I(x_1, x_2, t)}{\partial t} = D^{(2,2)} I(x_1, x_2, t) = \nabla I_{x_1} + \nabla I_{x_2} \tag{7}$$

where ∇I_{x_1} and ∇I_{x_2} denote the second-order derivatives of $I(x_1,x_2,t)$ in the x_1 and x_2 direction, respectively, and $D^{(2,2)}$ the second-order differential operator. The normalized Gaussian kernel is defined as

$$G(\boldsymbol{x}, t) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{\boldsymbol{x} \cdot \boldsymbol{x}}{2t}\right) \tag{8}$$

with $\boldsymbol{x} = (x_1, x_2)^{\mathrm{T}}$ and t denoting the variance (width) of the kernel. The integral $\int G(\boldsymbol{x}; t)\mathrm{d}\boldsymbol{x} = 1$, which means that convolving a signal with the Gaussian kernel, $I{*}G$, does not effect its average intensity level. The generalization of the linear scale space to $d > 2$ dimensions, $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^{\mathrm{T}}$, is straightforward.

Taylor Expansion Features

When $I(\boldsymbol{x})$ is a continuous, analytical function and all partial derivatives with respect to $\boldsymbol{x}$ are defined, it can be approximated in a neighbourhood around $\boldsymbol{x}0$ by a multidimensional Taylor expansion:[28]

$$I(\boldsymbol{x} - \boldsymbol{x}0) \cong I(\boldsymbol{x}0) + \sum_{\boldsymbol{h} \in \Lambda} D^{\boldsymbol{h}} I(\boldsymbol{x}0) \frac{(\boldsymbol{x} - \boldsymbol{x}0)^{\boldsymbol{h}}}{\boldsymbol{h}!} \tag{9}$$

with $\boldsymbol{h}$ an element in Λ, a set of multi-indices, and $D^{\boldsymbol{h}}$ the vectorial derivative with respect to the set $\{h_1, \ldots, h_d\}$, e.g., $\boldsymbol{h} = (1, 0, \ldots, 0)^T$ denotes the first-order derivative with respect to x_1. The notion $(\boldsymbol{x} - \boldsymbol{x}0)^{\boldsymbol{h}} = (x_1 - x0_1)^{h_1} \ldots (x_d - x0_d)^{h_d}$ and $\boldsymbol{h}! = h_1! \ldots h_d!$ From Taylor's theorem it follows that when the number of terms goes towards infinity the approximation, equation (9), becomes exact, given the assumptions of continuity and differentiability of I in $\boldsymbol{x}0$. Consequently, the manifold $I(\boldsymbol{x})$ can in the limit be characterized in the point $\boldsymbol{x}0$. This property of Taylor's expansion implies that optimal minimal error rate segmentation of an image can in theory be based on the complete set of vectorial derivatives $D^{\boldsymbol{h}}I(x)$, $\boldsymbol{h} \in \Lambda$, $\boldsymbol{x} \in X$. (It is well known that in many cases a Taylor expansion is not the most compact polynomial approximation of a differentiable function, for a discussion see, for example, Ralston.[29])

Let us define the complete set of derivatives $\boldsymbol{I}(\boldsymbol{x}) \equiv \bigcup_{\boldsymbol{h} \in \Lambda} D^{\boldsymbol{h}}I(\boldsymbol{x})$, the so-called *N*-jet,[24] and assume that the image element is classified by Bayes' rule:

$$P(\omega_j|\boldsymbol{I}(\boldsymbol{x})) = \frac{P(\omega_j)p(\boldsymbol{I}(\boldsymbol{x})|\omega_j)}{\sum_i P(\omega_i)p(\boldsymbol{I}(\boldsymbol{x})|\omega_i)} \tag{10}$$

The minimal error rate that can be obtained follows from equation (4):

$$\varepsilon^* = 1 - \sum_{j=1}^{c} P(\omega_j)P(\boldsymbol{I}(\boldsymbol{x}) \in R_j|\omega_j) \tag{11}$$

with $R_j = \{\boldsymbol{I}(\boldsymbol{x}) \mid P(\omega_j)\,p(\boldsymbol{I}(\boldsymbol{x}) \mid \omega_j) > P(\omega_i)\,p(\boldsymbol{I}(\boldsymbol{x}) \mid \omega_i), \forall i \neq j\}$. Hence, ε^* is the minimal error rate that can be obtained with the set of derivative features specified by the dimensionality of $\boldsymbol{\Lambda}$, i.e., the number of derivatives included in the feature vector. From the definition of R_j and the winner-takes-all rule (2) it follows that the prior and (overlapping) class-conditional distributions jointly determine the error rate that is obtained. As a consequence, the classification result with the minimal error rate ε^* can only be obtained when $P(\omega_j)$, $j = 1, \ldots, c$, constitute the probabilities that the patterns belong to the c different classes.

Assume that the prior distribution of the c classes in the training set is given by $P(\omega_j)$, $j = 1, \ldots, c$. The resulting classifier partitions the feature space into the disjoint regions, $R_1, \ldots, R_c$, as defined in equation (3). Assume further that in an actual image to be segmented, I, the prior distribution of occurrence of the different segments, is given by $P_I(\omega_j)$. If the observed prior distribution in the image I differs from that in the training set, $\exists j \in \{1, \ldots, c\}$, $P(\omega_j) \neq P_I(\omega_j)$, we can show that segmentation of I using Bayes rule, equation (10), results in an error rate $\varepsilon(P_I(\omega_j))$ that is always larger than or, at best, equal to the optimal error rate ε^*.

Theorem 1. The realized probability of error is always larger than or equal to the minimal error rate, $\varepsilon(P_I(\omega_1), \ldots, P_I(\omega_c)) \geq \varepsilon^*$, *in the case of overlapping class-conditional distributions,* $p(\boldsymbol{I}(\boldsymbol{x}) \mid \omega_j)$, $j = 1, \ldots, c$.

Proof. It follows from Duda and Hart[22] that the integral

$$\begin{aligned}\varepsilon^* &= \int_{\Re^d} P(error|\boldsymbol{I}(\boldsymbol{x}))p(\boldsymbol{I}(\boldsymbol{x}))\mathrm{d}\boldsymbol{x} \\ &= 1 - \sum_{j=1}^{c} P(\omega_j)P(\boldsymbol{I}(\boldsymbol{x}) \in R_j|\omega_j)\end{aligned} \tag{12}$$

should be as small as possible for every $\boldsymbol{x}$ which implies the use of the winner-takes-all rule, equation (2), i.e., the class label with the maximal posterior probability, $P(\omega_j \mid \boldsymbol{I}(\boldsymbol{x}))$, should always be assigned to $\boldsymbol{I}(\boldsymbol{x})$. Now, from McMichael[30] follows the exact correction for the novel prior probability by the formula

$$P_I(\omega_j|\boldsymbol{I}(\boldsymbol{x})) = \frac{\dfrac{P_I(\omega_j)}{P(\omega_j)}P(\omega_j|\boldsymbol{I}(\boldsymbol{x}))}{\dfrac{P_I(\omega_j)}{P(\omega_j)}P(\omega_j|\boldsymbol{I}(\boldsymbol{x})) + \dfrac{P_I(\overline{\omega}_j)}{P(\overline{\omega}_j)}P(\overline{\omega}_j|\boldsymbol{I}(\boldsymbol{x}))} \tag{13}$$

with $P(\overline{\omega}_j) = \Sigma_{i \neq j} P(\omega_i)$ and $P(\overline{\omega}_j|\boldsymbol{I}(\boldsymbol{x})) = \Sigma_{i \neq j} P(\omega_i|\boldsymbol{I}(\boldsymbol{x}))$. Classification of image elements by applying equation (2) to $P_I(\omega_j \mid \boldsymbol{I}(x0))$ gives the optimal segmentation result. In the case of overlapping class-conditional distributions, $P(\omega_j \mid \boldsymbol{I}(\boldsymbol{x})) \neq P_I(\omega_j \mid \boldsymbol{I}(\boldsymbol{x}))$, when $\exists j$: $P(\omega_j) \neq P_I(\omega_j)$. Consequently, $R_{I,j} \neq R_j$, with $R_{I,j} = \{\boldsymbol{I}(\boldsymbol{x}) \mid P_I(\omega_j \mid \boldsymbol{I}(\boldsymbol{x})) > P_I(\omega_i \mid \boldsymbol{I}(\boldsymbol{x})), \forall i \neq j\}$. Thus, for the error rate it holds that

$$\varepsilon(P_I(\omega_j)) = 1 - \sum_{j=1}^{c} P_I(\omega_j)P(\boldsymbol{I}(\boldsymbol{x}) \in R_j|\omega_j) \geq \varepsilon^* \tag{14}$$

with the minimal error rate being

$$\varepsilon^* = 1 - \sum_{j=1}^{c} P_I(\omega_j)P(\boldsymbol{I}(\boldsymbol{x}) \in R_{I,j}|\omega_j) \tag{15}$$

□

A direct consequence of Theorem 1 is that regardless of which features are provided as input to the Bayes' classifier, when the prior class distribution of the patterns differs from that of the training set used to build the classifier, the classification result will have an inferior overall error rate (in the case of overlapping class-conditional distributions and $P(\omega_j) > 0$, $P_I(\omega_j) > 0$, $j = 1, \ldots, c$). It is, however, possible to correct the classifier for a changed, prior distribution by means of the formula in equation (13).

We can furthermore prove the following lemma regarding the class-conditional error rate:

Lemma 2. The class-conditional probability of error ε_{ω_j} *is unchanged for any prior class distribution,* $P_I(\omega_j)$.

Proof. From the definition of the overall error rate

$$1-\varepsilon^{*}=\sum_{j=1}^{c}\int_{R_j}p(\boldsymbol{I}(\boldsymbol{x})|\omega_j)P_I(\omega_j)\mathrm{d}\boldsymbol{x} \tag{16}$$

follows the class-conditional error rate ε_{ω_j}

$$1-\varepsilon_{\omega_j}=\frac{\int_{R_j}p(\boldsymbol{I}(\boldsymbol{x})|\omega_j)P_I(\omega_j)\mathrm{d}\boldsymbol{x}}{\int_{\Re^d}p(\boldsymbol{I}(\boldsymbol{x})|\omega_j)P_I(\omega_j)\mathrm{d}\boldsymbol{x}} \tag{17}$$

Using Bayes' rule, the denominator may be rewritten as

$$\int_{\Re^d}p(\boldsymbol{I}(\boldsymbol{x})|\omega_j)P_I(\omega_j)\mathrm{d}\boldsymbol{x}=\int_{\Re^d}p(\omega_j|\boldsymbol{I}(\boldsymbol{x}))P_I(\boldsymbol{I}(\boldsymbol{x}))\mathrm{d}\boldsymbol{x} \tag{18}$$

which, according to the conditioning property, yields $P_I(\omega_j)$. So

$$1-\varepsilon_{\omega_j}=\frac{\int_{R_j}p(\boldsymbol{I}(\boldsymbol{x})|\omega_j)P_I(\omega_j)\mathrm{d}\boldsymbol{x}}{P_I(\omega_j)}=\int_{R_j}p(\boldsymbol{I}(\boldsymbol{x})|\omega_j)\mathrm{d}\boldsymbol{x} \tag{19}$$

which, for a given classifier, $R_1, \ldots, R_c$, is independent from the prior probability distribution in a particular image I, $P_I(\omega_j)$, $j=1, \ldots, c$. □

This lemma shows that the error rate per class remains constant irrespective of the actual class distribution in an image $P_I(\omega_j)$.

Segmentation under Zooming

We will now investigate the effect of changing the field of view in an image. First, we need to define the magnification function, M, which is responsible for zooming the image, M: $I(\boldsymbol{x})\times\Re^{+}\rightarrow I'(\boldsymbol{x})$:

$$I'(\boldsymbol{x})=M(I(\boldsymbol{x}),k),\ \ x_i\in X_i \tag{20}$$

where $M(I(\boldsymbol{x}),1)=I(\boldsymbol{x})$. Realizing that the magnification function will in most cases lead to a different prior distribution of the segments, $P_{M(I,k)}(\omega_j)$, it is clear that zooming will, in general, lead to an inferior overall error rate, unless the posterior probabilities are corrected for the novel prior probability distribution. We formulate two corollaries:

Corollary 3. Pixel-based segmentation algorithms based on statistical pattern classification give a higher overall error rate $\varepsilon(P_I(\omega_j))$ than the minimal error rate ε^*, when the prior probability $P_I(\omega_j)$ of one or more of the segments differs between training and test sets, e.g., as a result of zooming.

Corollary 4. Pixel-based segmentation algorithms based on statistical pattern classification give the same class-conditional error rate error ε_{ω_j} *for any prior probability distribution, $P_I(\omega_j)$, $j=1, \ldots, c$.*

As a consequence, when the accumulated size of a segment increases or decreases, the overall error rate becomes inferior whereas the relative error per type of segment remains constant. It should be noticed that it is not the magnification operation itself that can cause the prior probabilities of the different segments to change. The prior probabilities change because an image is a discrete signal with a fixed sampling scheme. Hence, zooming in on, for example, a certain specific texture implies that the other textures comprise a smaller part of the image content (the field of view). It follows directly from Lemma 2 that the class-conditional error rates ε_{ω_j}, $j=1, \ldots, c$, remain constant irrespective of which magnification factor k is chosen.

Scale Selection in the Discrete Scale Space

Discrete Scale Space

In digital image processing, the image $I(\boldsymbol{x})$ is a discrete signal with a finite number of sample points. Computation of derivatives of a discretely sampled signal is an ill-posed problem. In the linear scale space, differentiation is performed by convolution with derivatives of the Gaussian kernel, which transforms differentiation into a well-posed problem.[24] It has also been shown that a convolution with Gaussian derivative kernels satisfies equation (7). Hence, regularized differentiation of a discrete image relies on the equivalence $D^{\boldsymbol{h}}I(\boldsymbol{x})*G(\boldsymbol{x};t)=I(\boldsymbol{x})*D^{\boldsymbol{h}}G(\boldsymbol{x};t)$, with $*$ indicating the convolution operation. The crux in the linear scale space lies in the commutative properties of these two steps because one can instead differentiate the (blurred) Gaussian kernel and subsequently perform the convolution with the image.

In the discrete case, we can rewrite the Taylor expansion, equation (9), as

$$I_t(\boldsymbol{x}-\boldsymbol{x}0)\cong I(\boldsymbol{x}0)+\sum_{\boldsymbol{h}\in\Lambda}(D^{\boldsymbol{h}}G(\boldsymbol{x}0;t)*I(\boldsymbol{x}0))\frac{(\boldsymbol{x}-\boldsymbol{x}0)^{\boldsymbol{h}}}{\boldsymbol{h}!} \tag{21}$$

in which the convolution with the Gaussian derivative operator facilitates the regularized differentiation of $I(\boldsymbol{x})$, for $\boldsymbol{x}=\boldsymbol{x}0$. Note that in the continuous case $I(\boldsymbol{x})*D^{h}G(\boldsymbol{x};t)\rightarrow D^{h}I(\boldsymbol{x})$ for $t\rightarrow 0^{+}$. We will now define the discrete equivalent of $\boldsymbol{I}(\boldsymbol{x})$, the complete set of derivatives computed for a set of different scales, $\boldsymbol{I}(\boldsymbol{x}0,\Sigma) \equiv \bigcup_{t\in\Sigma} \bigcup_{\boldsymbol{h}\in\Lambda} D^{\boldsymbol{h}}G(\boldsymbol{x}0;t) * I(\boldsymbol{x}0)$, with $\Sigma = (t_1, \ldots, t_s)^T$ denoting the set of scales at which the derivatives are computed. This set of derivatives is called the discrete N-jet.[24] The discrete Taylor expansion, equation (21), differs from its continuous counterpart, equation (9), because of the differential operator which is used to compute the regularized derivatives pertaining to the discrete image $I(\boldsymbol{x})$. As a result, there exists a minimal scale at which images can be segmented. Based on an analysis of the density of local extrema in the continuous and the discrete scale spaces, Lindeberg computed the minimal scale at which these two density functions correspond.[31] His analysis indicates that for a value of t below 0.5848^2, the continuous analysis is not a valid approximation of its discrete counterpart. Moreover, a reliable and stable computation of derivatives requires even more blurring; the higher the derivative, the more regularization is required.[32]

Within the framework of the linear scale space theory, scale itself is treated as a free parameter that is varied across all possible scales.[31] The scale at which a particular scale space feature detector (e.g., a junction detector) gives the maximal response, is considered the natural scale of the located feature. However, whereas the method for scale selection proposed by Lindeberg[31] and elaborated in Lindeberg[33] works well for noise-free, sharp images, his experiments also show that either a slight blur or a perturbation by noise will result in a different natural scale being selected.[33] This is caused by the trade-off between *bias* and *variance*, which is implicitly made during scale selection. Blurring with a *wide* Gaussian kernel—the generating function in the linear scale space—results in a robust detection result, which is insensitive to the random components in the high-frequency noise in the image. However, the location accuracy of a feature detector operating at a coarse scale is poorer than the accuracy obtained by the same detector applied at a finer scale.[31] Although much blurring suppresses high-frequency noise, the finer-scaled edges in the image migrate—the extent of the migration increases with the width of the kernel used for blurring the image. This migration is essentially a *location bias*. It is clear that in the presence of noise the choice of an appropriate scale in a segmentation approach enforces a trade-off between location bias and variance. Recognizing that scale selection remains an ill-posed problem in signal-driven segmentation, we propose to let a Bayesian classifier perform scale selection by optimization of an error criterion.

Scale Selection: Balancing Between Bias and Variance

Computation of the (derivative) features in the linear scale space essentially consists of two steps: blurring (regularization) with the Gaussian kernel followed by differentiation of the image. In the sequel, we will study the effect of blurring the image in the presence of additive, Gaussian noise.

Define the 'noisy' image $I_e(\boldsymbol{x})$ as

$$I_e(\boldsymbol{x}) = I(\boldsymbol{x}) + e(\boldsymbol{x};\sigma^2) \tag{22}$$

with $e(\boldsymbol{x};\sigma^2)$ an additive, Gaussian-distributed noise term, $e(\boldsymbol{x};\sigma^2) \sim U(\mu_e, \sigma^2)$, with a zero mean μ_e and a variance σ^2. Blurring the image $I_e(\boldsymbol{x})$ with the Gaussian kernel G with the scale parameter t yields

$$I_e(\boldsymbol{x}) * G(\boldsymbol{x};t) = \int_{\boldsymbol{x}'\in X} I_e(\boldsymbol{x}-\boldsymbol{x}')\cdot G(\boldsymbol{x}';t)\mathrm{d}\boldsymbol{x}' \tag{23}$$

which equals

$$\int_{\boldsymbol{x}'\in X} (I(\boldsymbol{x}-\boldsymbol{x}') + e(\boldsymbol{x}-\boldsymbol{x}';\sigma^2))\cdot G(\boldsymbol{x}';t)\mathrm{d}\boldsymbol{x}' \tag{24}$$

and

$$\int_{\boldsymbol{x}'\in X} I(\boldsymbol{x}-\boldsymbol{x}')\cdot G(\boldsymbol{x}';t) + e(\boldsymbol{x}-\boldsymbol{x}';\sigma^2)\cdot G(\boldsymbol{x}';t)\mathrm{d}\boldsymbol{x}' \tag{25}$$

This convolution integral partitions into

$$I_e(x) * G(\boldsymbol{x};t) = \underbrace{\int_{\boldsymbol{x}'\in X} I(\boldsymbol{x}-\boldsymbol{x}')\cdot G(\boldsymbol{x}';t)\mathrm{d}\boldsymbol{x}'}_{\text{Bias term}} + \underbrace{\int_{\boldsymbol{x}'\in X} e(\boldsymbol{x}-\boldsymbol{x}';\sigma^2)\cdot G(\boldsymbol{x}';t)\mathrm{d}\boldsymbol{x}'}_{\text{Variance term}} \tag{26}$$

The bias term represents the part of the original (noise-free) image that is retained, i.e., the image after high-frequentcy details have been removed. The variance term indicates the result of blurring away the additive noise. The following property holds:

Proposition 5 The variance term of $I_e(\boldsymbol{x}) * G(\boldsymbol{x};t)$ approaches μ_e for $t\to\infty$. When $\mu_e = 0$, the variance term vanishes.

Proof. A convolution with the Gaussian function can in the Fourier space be written as $\Im\{e(\boldsymbol{x};\sigma^2) * G(\boldsymbol{x};t)\} = \Im\{e(\boldsymbol{x};\sigma^2)\}\cdot\Im\{G(\boldsymbol{x};t)\}$. The Fourier transform of the

Gaussian function $\Im\{G(x;t)\}$ is also a Gaussian function. For $t \to \infty$, $\Im\{G(x;t)\}$ becomes a Dirac pulse, for a two-dimensional image $\delta(u,v)$. As $\delta(0,0)=1$ and $\delta(u,v)=0$, $\forall(u,v)\neq(0,0)$, solely the mean μ_e is retained from $\Im\{e(x;\sigma^2)\}\cdot\delta(u,v)=X_i\ \mu_e$, with X_i, the number of pixels in a row in the quadratic image $I(\boldsymbol{x})$. It follows directly that when $\mu_e=0$ the variance term vanishes. This result generalizes by induction to d-dimensional images. □

Blurring away noise has a price, namely an increased (location) bias:

$$I(\boldsymbol{x}) - \int_{\boldsymbol{x}'\in X} I(\boldsymbol{x}-\boldsymbol{x}')\cdot G(\boldsymbol{x}';t)\mathrm{d}\boldsymbol{x}' \qquad (27)$$

It is a consequence of the linear diffusion equation (7) that blurring with the Gaussian kernel introduces a location bias; the largest bias occurs around the edges in the image (so-called edge migration) whereas large homogeneous areas remain largely unchanged.

Classification in the Presence of Noise

Until now, we have only examined the influence of additive noise and blurring on the distorted image $I_e(\boldsymbol{x})$. In the sequel, we will incorporate these aspects into the statistical model of the image, which forms the basis of minimal error rate segmentation as sketched above. Let $z=\boldsymbol{I}(x)$ denote the set of vectorial-derivative features derived from the image. When we assume that the additive noise term is independent of the coordinate in the image, the resulting marginal probability density function can, in the presence of noise, be written as a convolution:[34]

$$p_e(\boldsymbol{z}) = \int_{\Re^d} p(\boldsymbol{z}-\boldsymbol{z}')p(e(\boldsymbol{z}';\sigma^2))\mathrm{d}\boldsymbol{z}' \qquad (28)$$

with $e(z';\sigma^2)$ indicating the noise added to the feature vector z'. The resulting marginal probability density function, $p_e(\boldsymbol{z})$, is wider than the noise-free density $p(\boldsymbol{z})$. Writing the marginal density as $p(\boldsymbol{z})=\Sigma_j\ p(\boldsymbol{z}\mid\omega_j)P(\omega_j)$, equation (28) can be rewritten as

$$p_e(\boldsymbol{z}) = \int_{\Re^d} \sum\nolimits_j p(\boldsymbol{z}-\boldsymbol{z}'|\omega_j)P(\omega_j)p(e(\boldsymbol{z}';\sigma^2))\mathrm{d}\boldsymbol{z}' \qquad (29)$$

which equals

$$p_e(\boldsymbol{z}) = \sum\nolimits_j P(\omega_j)\int_{\Re^d} p(\boldsymbol{z}-\boldsymbol{z}'|\omega_j)p(e(\boldsymbol{z}';\sigma^2))\mathrm{d}\boldsymbol{z}' \qquad (30)$$

It is clear that the additive noise term entails a convolution of each class-conditional distribution with the probability density function $p(e(z';\sigma^2))$. Consequently, the class-conditional distributions will have a larger overlap in the presence of noise which gives a poorer classification result.

An analytical study of the effect the location bias has on the (optimal) error rate ε^*, presupposes that the true (noise free) function $I(\boldsymbol{x})$ is known. This is generally not the case in practical image processing. We propose to use the overall error rate, $\varepsilon(\Sigma)$, to determine the optimal blurring that results in the best possible segmentation result, where the error rate $\varepsilon(\Sigma)$ is considered a function of the sampling scheme Σ in the linear scale space.

A Statistical Approach to Scale Selection

Segmentation approaches based on features from the (linear) scale space require a mechanism for scale selection, i.e., the dimension and entries in the sampling scheme, Σ. Neglecting this issue will lead to inferior segmentation results as was shown by previous experiments.[9,21] For most practical segmentation tasks, a specific range of meaningful scales Σ can be specified. This is especially the case in tomographic medical imaging (MRI, CT) where the absolute size of each voxel is completely determined by the acquisition protocol. Scale selection requires the following two choices: the limits of the sampled interval in the scale space, $\min(\Sigma)$ and $\max(\Sigma)$, and the sampling density in the scale space, $\mathrm{card}(\Sigma)$. (Note that the scale space is normally logarithmically sampled.[31] If the features do only consist of images blurred at different scales, an obvious choice for the minimal scale is the so-called 'inner scale' of the image, i.e., no blurring takes place, so $\min(\Sigma)=0$. When derivative features are also included in the feature set z, a larger minimal scale should be chosen, $\min(\Sigma)>0$. With respect to the maximal scale, the choice is determined by the maximal size of the segments one wants to detect, the noise level in the image material and the scale of the underlying structures that are being imaged. We propose the following algorithm for selecting the optimal sampling of the scale space:

Scale selection algorithm: scale-invariant segmentation

1. Compose a set of training images with a uniform prior class distribution.
2. Choose the minimal scale $\min(\Sigma)$. If only blurred features are included, $\min(\Sigma)$ can be the inner scale

of the image. Otherwise, the minimal scale needs to be varied also, $\min(\Sigma) > 0$.

3. Choose a maximal scale, $\max(\Sigma)$, and a (logarithmic) sampling scheme Σ.
4. Train a statistical classifier with the chosen features from the scale space, $\boldsymbol{I(x,\Sigma)}$.
5. Compute the error rate $\varepsilon(\Sigma)$ of the classifier on a test set that is representative for the desired application. If sufficient, stop; otherwise, choose a new (minimal and) maximal scale, and sampling scheme, go to step 4.

Experiments

We performed a number of experiments with the algorithm for scale selection defined above. First, a set of quality measures is defined that makes it possible to study the effect of varying the sampling scheme in the scale space on the accuracy and the homogeneity of the obtained segmentation results. Subsequently, a classifier is chosen and used in an experiment with synthetic data to illustrate the theoretical results. Finally, a number of experiments with segmentation of dynamic contrast-enhanced MR images is performed.

Spatial Quality Measures

The performance of a segmentation algorithm can be assessed with true class labelling by computing a contingency table A, with $a_{i,j}$ the number of voxels that are classified into class i while belonging to class j. The correctness φ, the fraction of voxels that is classified correctly, and κ, the same fraction corrected with respect to the prior distributions, are derived from the contingency table[35] as follows:

$$\varphi = 1 - \varepsilon = \frac{\sum_{i=1}^{c} a_{i,i}}{\sum_{i=1}^{c}\sum_{j=1}^{c} a_{i,j}} \tag{31}$$

$$\kappa = \frac{\varphi - q}{1 - q}, \quad q = \frac{\sum_{i=1}^{c} \left(\sum_{g=1}^{c} a_{i,g} \cdot \sum_{g=1}^{c} a_{g,i} \right)}{\left(\sum_{i=1}^{c}\sum_{j=1}^{c} a_{i,j} \right)^2} \tag{32}$$

These performance measures do not take into account whether misclassified voxels are scattered all over the image or form one or a few connected clusters. To assess the effect of including images at different scales in the training set, we modified two existing spatial quality measures[36] by making them isotropic. Both spatial quality measures are based on the local *entropy* of the class labels in the labelled image and measure essentially spatial scatter and dispersion in a neighbourhood $W(\boldsymbol{x})$. Define the entropy image $H(\boldsymbol{x})$:

$$H(\boldsymbol{x}) = -\sum_{j=1}^{c} P_j \cdot \frac{\ln(P_j)}{\ln(c)} \quad \text{with } P_j = \frac{\sum_{x \in W(x)} S(\boldsymbol{x}) = j}{\mathrm{card}(W(\boldsymbol{x}))} \tag{33}$$

where c denotes the number of classes, $S(\boldsymbol{x})$ the label assigned to voxel $\boldsymbol{x}$, card$(\cdot)$ the cardinality function (number of elements) and $W(\boldsymbol{x})$ a circular window. Define the class-conditional *confidence* θ_j as

$$\theta_j \equiv 1 - \frac{\sum_{x \in \omega_j} H(\boldsymbol{x})}{\mathrm{card}(\omega_j)} \tag{34}$$

with ω_j denoting the set of voxels truly belonging to class j. The size of the circular neighbourhood $W(\boldsymbol{x})$ determines the scale at which the confidence is computed.

The class-conditional confidence is the estimated mean of the local entropy in the labelled image. We define also a dispersion measure, the class-conditional uniformity, to capture the variation around this mean. Define the class-conditional *uniformity* γ_j as

$$\gamma_j \equiv 1 - \sum_{i=1}^{q} P_{i,j} \cdot \frac{\ln(P_{i,j})}{\ln(q)} \quad \text{with } P_{i,j} = \frac{\mathrm{card}(\lambda_i)}{\mathrm{card}(\omega_j)} \tag{35}$$

where $\lambda_i = \{\boldsymbol{x} \mid i \leq H(\boldsymbol{x}) < i+1\}$ denotes the set of voxels $\boldsymbol{x} \in X$ that have the entropy level i. The class-conditional *confidence* θ_j expresses the average entropy in a neighbourhood $W(\boldsymbol{x})$ whereas the *uniformity* γ_j is a measure for the dispersion around θ_j. Both spatial quality measures are computed for the labelled image $S(\boldsymbol{x})$.

Choice of Classifier

The optimal classifier with the error rate ε^* generally entails a non-linear partitioning of the feature space in regions R_1, ..., R_c. Neural networks with one hidden layer have been shown to approach the minimal error rate classifier when the number of hidden nodes and the size of the learning set both go towards infinity.[37] It may even approximate a quadratic discriminant exactly after training with a gradient-descent learning algorithm.[38] Finally, a feed-forward neural network with one hidden layer is capable of approximating any continuous discriminant function.[39,40]

In our application, segmentation of dynamic MR images, the type of the underlying feature distributions is unknown so we decided to use a multi-layer feed-forward neural network as classifier. For neural networks to perform best in classification tasks, each output node should represent a specific class.[41] As it has been shown that feed-forward neural networks approach the Bayes minimal error rate classifier, the elements of the output vector $\boldsymbol{o}_i$ can be interpreted as posterior probabilities, $o_j \approx P(\omega_j | \boldsymbol{z})$, $j = 1, \ldots, c$.

Experimental Set-Up

A set of experiments with perfusion MR images of patients with bone tumours (Ewing's sarcoma) was conducted. The perfusion of blood in the tissue under study is assessed by continuously acquiring a sequence of MR images while a bolus of MR contrast tracer (Gd-DTPA) is given intravenously.[42] The primary goal was to develop a scale-invariant, signal-driven segmentation approach based on a statistical classifier. It was shown[9,21,42] how features derived from pharmacokinetic functions, fitted onto the dynamic MR signal, can be used for segmentation into viable tumour, non-viable (necrotic) tumour and healthy (normal) tissue. However, this segmentation approach requires excessive computation since a pharmacokinetic model has to be fitted to the dynamic MR signal associated with each voxel. Instead, we developed a segmentation approach based on the normalized dynamic MR signal, $I(\boldsymbol{x},\tau)$, given by the dynamic sampling scheme $\tau = 1, \ldots, T$.

We propose a signal-driven segmentation algorithm, which will be based on a sample obtained from the linear scale space:

$$I(\boldsymbol{x},\tau,t) = I(\boldsymbol{x},\tau) * G(\boldsymbol{x};t), \quad t \in \Sigma \tag{36}$$

with $G(x;t)$ denoting the Gaussian kernel. Previous experiments[9,21] have indicated that the inclusion of derivative features $I(\boldsymbol{x},\tau)*D^hG(\boldsymbol{x};t)$ does not improve the segmentation result for bone tumours. Consequently, such derivative features were omitted in the experiments reported here. Instead, it will be shown how dynamic MR images can be segmented from the shape of the dynamic MR signal (see Figure 1).

The dynamic MR signal is affected by both a random

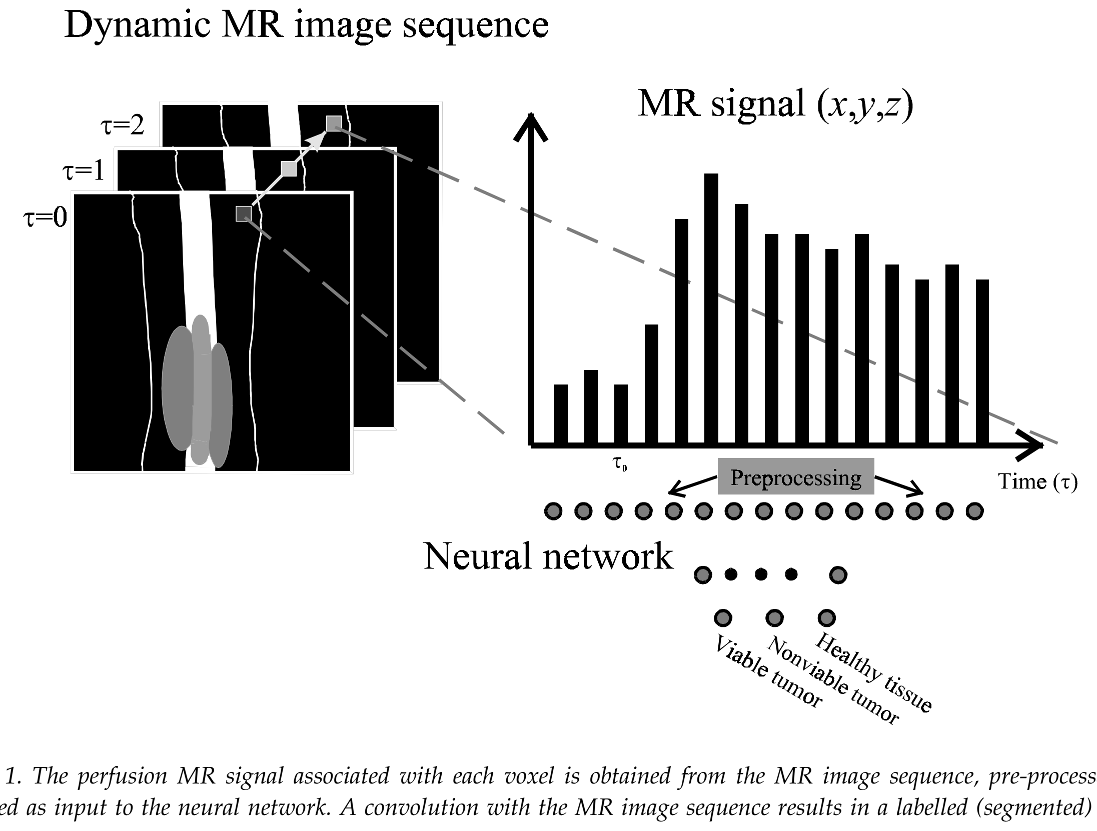


Figure 1. The perfusion MR signal associated with each voxel is obtained from the MR image sequence, pre-processed and provided as input to the neural network. A convolution with the MR image sequence results in a labelled (segmented) image.

and a systematic distortion: the noise introduced by the MR device and the MR signal fluctuations caused by the heartbeat. The fluctuations associated with the heartbeat are caused by a combination of uneven mixing of the bolus in the blood, the pumping of the left ventricle and the flow sensitivity attributed to the chosen MR sequence. Although a simple low-pass filtering may eliminate a large part of the high-frequency noise, this technique is not suitable in our application because important high frequencies that characterize the uptake speed of the contrast tracer would also be heavily damped. Morphological operators, on the other hand, retain the large edges. The morphological min–max filter,[43] which we used for preprocessing, is defined as

$$I_{\text{filter}}(\tau, w) = \frac{\max\limits_{y \in \boldsymbol{b}}\left(y = \min\limits_{\tau \in \boldsymbol{b}}(I(\tau))\right) + \min\limits_{y \in \boldsymbol{b}}\left(y = \max\limits_{\tau \in \boldsymbol{b}}(I(\tau))\right)}{2} \qquad (37)$$

with $b(\tau) = \{\tau - (w-1)/2,\ \tau + (w-1)/2\}$ and $I(\tau)$ denoting the intensity associated with voxel $I(\boldsymbol{x}, \tau)$.

Generally, the strength of signals obtained from an MR scanner depends on several factors such as the tumour location, the weight of the patient and the relaxivity of the surrounding tissue. The most important factor determining the signal amplitude and level is, however, the affine scaling that is performed by the software on the MR scanner. Besides this scaling, there is also an intensity offset that differs between the various scans as well as within one scan. To correct for these differences, an offset estimation is made by calculating the mean of the sample points before τ_0 (when injection with paramagnetic tracer is started). The signal $\hat{I}(\tau)$ is normalized according to

$$\hat{I}(\tau) = I(\tau) \cdot RS - I_{\text{offset}}, \quad \tau \in T \qquad (38)$$

with RS denoting the scale factor, obtained from the MR scanner and I_{offset} the intensity offset.

Before our approach to scale selection can be evaluated on the dynamic MR images, the correct segmentation result, i.e. the correct class membership of each voxel, is required. In some applications, the correct segmentation result has to be specified by a mask, annotated by a human expert.[8] However, such manually drawn annotation masks are subject to intra- and inter-observer variation. Moreover, in our application the desired segmentation result cannot be annotated in the MR image sequence as complex characteristics of the dynamic MR signal are indicative for the correct class membership of each voxel. Instead, we benefit from the fact that for bone sarcoma the postoperative histological specimen is regarded as an objective gold standard when it comes to tissue classification. In our application, segmentation masks are obtained by matching postoperative specimens with the MR images according to a method developed in Egmont-Petersen *et al.*[42] and applied in Frangi *et al.*[9] and Egmont-Petersen *et al.*[21] The differences in scale, orientation and position between the MR section and the histological macroslice are computed using a method based on the principal axes of the coordinate sets obtained by sampling the contours;[44] for details see Egmont-Petersen *et al.*[21] The matched histological macroslices constitute masks that indicate the true class membership of each voxel in the MR images.

Experiment with Synthetic Data. The purpose of the first experiment was to verify the theoretical results (Theorem 1 and Lemma 2) regarding the effect of magnification on the change in the overall and class-conditional error rates. Three synthetic MR image sequences, each with two segments representing viable tumour and healthy tissue, were generated. Each MR image sequence consisted of 25 dynamic images with a resolution of 256 × 256 pixels. In each sequence, a circle with a specific radius, respectively 256 pixels, 81 pixels and 26 pixels, indicates the tissue 'viable tumour', which resulted in the following prior distributions: $P_{I1}(\text{viable}) = 0.79$, $P_{I2}(\text{viable}) = 0.079$ and $P_{I3}(\text{viable}) = 0.008$. The segment 'viable tumour' was based on the MR signal extracted from a central part of an area with viable tumour in a typical patient, as indicated by the registered mask. A dynamic MR signal representative for 'healthy tissue' was extracted in a similar way. Gaussian-distributed noise with a zero mean and the variances $\sigma^2(\text{healthy}) = 5$, $\sigma^2(\text{viable}) = 10$ was added to the respective signals, and the three synthetic image sequences were combined according to the three masks. A neural network was subsequently trained during 5000 cycles with standard back-propagation[45] to segment the images, based on a training set with 10,000 patterns where each of the two classes had the same prior probability, $P(\text{viable}) = P(\text{healthy})$. The feed-forward network had 25 input nodes, eight hidden nodes and one output node, associated with the class *viable tumour*. In the training set, the desired output was set to 1 for viable tumour and 0 for healthy tissue. The neural network was subsequently applied to the three synthetic image sequences (see Figure 2).

The performance of the trained neural network was analysed in two ways on the three resulting output

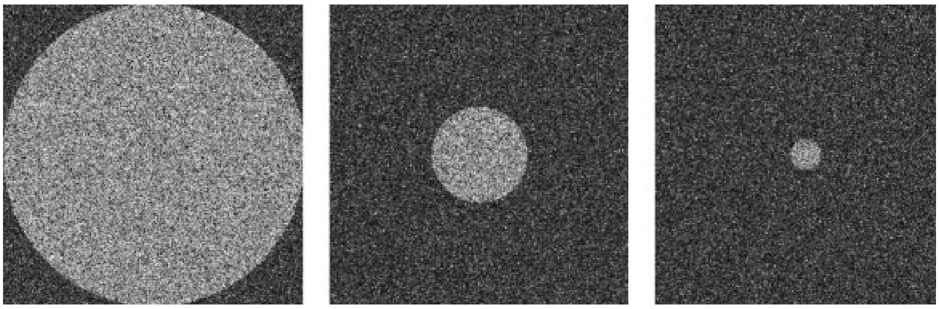


Figure 2. The segmented synthetic MR images contain different fractions of voxels that belong to the class viable tumour (indicated with a bright colour). The radii of the three viable segments are 256, 81 and 26 pixels, respectively.

images: according to an *uncorrected* segmentation scheme (the output image obtained by convolution with the neural network was classified using the winner-takes-all rule, equation (2)); and according to a *corrected* scheme: each of the three output images was post-processed according to the rule of McMichael,[30] equation (13), with the correct prior probability, before the winner-takes-all rule was applied.

The results are shown in Table 1. This experiment confirms that the class-conditional performance is unaffected by the change in prior probability, P_I(viable), in the uncorrected situation. To optimize the overall error rate φ, the correction formula, equation (13), should be applied, whereas constant class-conditional error rates can be obtained from a uniform prior (in the training set), thereby leaving the output of the classifier uncorrected.

First Experiment with MR Image Sequence.

For all further experiments, we constructed a pattern set consisting of data chosen at random from six different patients with bone tumours using the masks as defined from the corresponding histological images.[9,21] The complete pattern set contained the dynamic MR signals of each individual voxel, and was subsequently split into a training and a test set consisting of 10,000 and 2500 voxels, respectively.

In an earlier pilot experiment, we had used a training set in which the prior probabilities for the three classes were as observed in the set of images. These experiments gave poor classification results as the networks were unable to classify correctly any dynamic MR signal pertaining to the most infrequent class, viable tumour. Therefore, in our training set the three classes had the same prior probability (a uniform prior). The test pattern set had prior probabilities $P_I(\omega_j)$, $j=1$, ..., c, as averaged over the whole MR image data set, because we eventually want to select a neural network that performs well on representative image material. The width of the min–max filter and the number of hidden nodes were both varied. The results on the test set, consisting of 2500 dynamic MR signals, are shown in Table 2.

The first experiment (top part of Table 2) indicates that eight hidden nodes result in a good performance on the test set, especially for $|b|>7$. We experimented further with this network topology while increasing the width $|b|$ of the min–max filter. A filter width of 15 appeared to perform best in combination with a

	Uncorrected			Corrected		
	Circle 256	Circle 81	Circle 26	Circle 256	Circle 81	Circle 26
$1-\varepsilon_{\text{Viable}}$	0.71	0.72	0.71	0.88	0.44	0.00
$1-\varepsilon_{\text{Healthy}}$	0.96	0.96	0.96	0.71	0.99	1.00
κ	0.32	0.13	0.05	0.26	0.10	0.01
φ	0.77	0.94	0.96	0.84	0.95	0.99

Table 1. Class-conditional correctness, kappa and overall correctness, for all three synthetic MR images with no correction, and corrected for the different prior distributions

Filter width		Number of hidden nodes					
		1	2	3	4	8	16
0	κ	0.3559	0.4589	0.4765	0.4864	0.4820	0.4906
	φ	0.6665	0.7522	0.7672	0.7711	0.7664	0.7743
5	κ	0.3549	0.4569	0.4665	0.4765	0.4995	0.4953
	φ	0.6657	0.7515	0.7573	0.7664	0.7818	0.7782
7	κ	0.2783	0.4567	0.4652	0.4706	0.4845	0.4920
	φ	0.5919	0.7459	0.7537	0.7577	0.7668	0.7739
11	κ	0.2794	0.4865	0.4921	0.4956	**0.5104**	**0.5159**
	φ	0.5955	0.7656	0.7727	0.7727	**0.7849**	**0.7889**
		Filter width					
	8 hidden nodes	13	15	17	21	25	29
	κ	0.5106	**0.5166**	0.5100	0.5081	0.5103	0.4984
	φ	0.7865	**0.7889**	0.7861	0.7873	0.7857	0.7802

Table 2. Correctness and kappa values obtained for different combinations of the filter width $|b|$ in equation (36), and network topology. All statistics were computed on a representative test set containing 2500 patterns. Bold face indicates well-performing configurations

neural network with eight hidden nodes. In the remaining experiments in this article, we have used these values for filter width and number of hidden nodes.

Experiment with Scale Selection Approach. In this experiment, we investigated our approach to scale selection and studied the effects of varying the sampling of the scale space on performance. As earlier experiments with the same image material indicated that derivative features do not contribute to the segmentation of the dynamic MR images of bone tumours,[9,21] we decided solely to use blurred versions of the dynamic MR images as extra input to the classifier. As blurring is a regularization operation, it is to be expected that the classifier that performed optimally in the first experiment will also perform well when the blurred image data are added as input. Of course, the signal intensity obtained from the original image sequence is highly correlated with the intensity obtained from the blurred versions of the same dynamic MR images.

The previous experiment confirmed the added value of the min–max filter for pre-processing the dynamic MR signal. This result may indicate that a spatial (grey-level) morphological smoothing operator such as an opening[46,47] can improve the segmentation result. Furthermore, it makes it possible to broaden the scope of our experimental evaluation of the approach to scale selection by experimenting also with feature images from the morphological scale space. Let $I(\boldsymbol{x})$ be a single MR image at time τ. The morphological opening is defined as

$$\mathrm{Opening}(I(\boldsymbol{x}),B) = \max_{\boldsymbol{x}'\in B}\,(\boldsymbol{x}' = \min_{\boldsymbol{x}\in B}\,(I(\boldsymbol{x}))\,) \qquad (39)$$

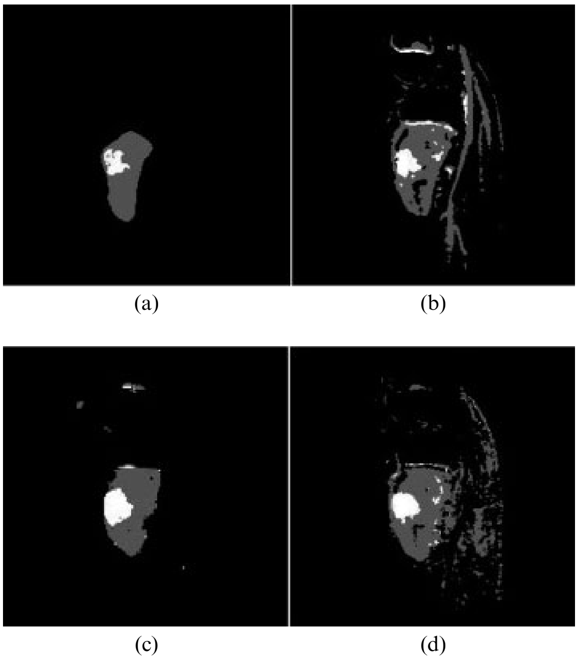
where B denotes the kernel, which is in this case an isotropic disc, $r = \mathrm{rad}(B)$.[46,47] The size of the kernel, measured in millimetres, must again be equal for all MR scans so the absolute scales are kept constant for all patients.

We used the 'inner scale' of the dynamic MR images in conjunction with either the image obtained from the linear or the morphological scale space, respectively. The maximal scale, max(Σ), was varied while studying the resulting performance of the trained neural networks on the representative test set. The kappa and correctness measures as well as the spatial quality measures were computed. Class-conditional confidence and uniformity were solely computed for the classes viable and non-viable tumour. Table 3 shows the width of the Gaussian and morphological kernels, measured in square millimetres.

When comparing features from the linear scale space, we provided as input to the neural classifier the two largest kernels ($t = 12$ and $15\,\mathrm{mm}^2$) resulting in

Kernel	σ^2	Kernel size (mm^2)				
		5	8	10	12	15
Gauss:	κ	0.5649	0.5811	0.5674	0.5874	**0.5964**
linear	φ	0.8772	0.8796	0.8788	0.8852	**0.8858**
scale	θ_1	0.7469	0.7398	0.7847	**0.8549**	0.7983
space	γ_1	0.2541	0.2493	0.3262	**0.4399**	0.2977
	θ_2	0.6422	0.7160	0.6986	**0.7466**	0.6993
	γ_2	0.2251	**0.3535**	0.2498	0.2780	0.2390
Disc:	κ	0.4670	0.4754	**0.5078**	0.4780	0.4805
morphological	φ	0.8434	0.8465	**0.8604**	0.8434	0.8478
scale	θ_1	0.6561	0.6426	**0.6953**	0.6404	**0.7067**
space	γ_1	0.2077	0.1944	0.2038	**0.2137**	0.2110
	θ_2	0.5287	**0.6446**	0.5365	0.6263	0.5565
	γ_2	0.1375	0.1829	0.1281	**0.2250**	0.1009

Table 3. Kappa and correctness measures, confidence and uniformity, as computed by applying the neural networks to the test set. The most suited kernel size for the linear scale space is 12–15 mm^2; for the morphological scale space it is 10 mm^2. Bold face indicates well-performing configurations



(a) (b)

(c) (d)

Figure 3. Segmentation results for one patient (Ewing's sarcoma present in the tibia), with the white regions denoting viable tumour, the grey regions non-viable tumour and the black regions background/healthy tissue: (a) the histological mask; (b) the result of a voxel-based classification (first experiment); (c) adding features from the linear scale space; (d) adding features from the morphological scale space.

the highest kappa and correctness measures. The local confidence θ_j (averaged over all six patients) was maximal for $t = 12\,\mathrm{mm}^2$. As shown by the resulting image of a patient in Figure 3(c), the large artery was correctly labelled as healthy tissue and the scatter in the image was reduced. For features obtained from the morphological scale space, the disc with a radius of 10 mm^2 resulted in the best performance as measured with the kappa and correctness measures. The spatial quality measures fluctuate as a function of the disc size but, more important, the spatial quality is poorer than obtained with features from the linear scale space. Consequently, the amount of scatter in the labelled image (e.g., Figure 3d) is higher than that obtained with features from the linear scale space.

This experiment indicates the feasibility of using the classification result obtained with a representative test set to choose the appropriate set of scales in the linear or morphological scale spaces. Another example can be seen in Figure 4, showing the segmentation results for a different patient.

The performance measures can be used to find the scale(s) that give the best segmentation result for the image material at hand (represented by the test set). It is clear that the set of scales Σ resulting in the maximal number of correctly segmented voxels does not have to coincide with the set of scales with the highest maximal confidence. Consequently, a trade-off may have to be made between performance and spatial coherence.

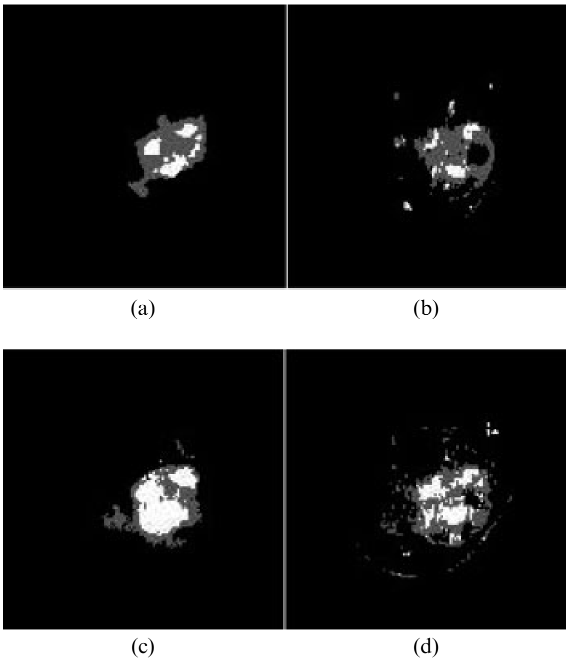


Figure 4. Segmentation results for another patient (Ewing's sarcoma), with the white regions denoting viable tumour, the grey regions non-viable tumour and the black regions background/healthy tissue: (a) the histological mask; (b) the result of a voxel-based classification (first experiment); (c) adding features from the linear scale space; (d) adding features from the morphological scale space.

Such performance trade-offs often occur in practice where an algorithm has to fulfil more, possibly contradictory requirements. For a discussion of this issue see, for example, Egmont-Petersen *et al.*[35] and Karthaus *et al.*[48]

The best neural network has been used to segment a 3D dataset consisting of eight MRI sections of a bone tumour located in the bone marrow of a femur (hip). The segmentation result was visualized (see Figure 5) by combining two volume rendering techniques: iso-surface rendering and maximal-intensity projection. Iso-surface rendering was used to indicate healthy bone marrow, which has a high intensity in T1-weighted MR images. The same technique was used to display the extent of the viable tumour remnants by thresholding the output of the neural network. Three maximal intensity projections of the image data, sagittal, transversal and coronal, depict the original MR imaging data.

Discussion

The experiments we performed with synthetic and real MR images support the theoretical results presented in this paper. It is possible to use a statistical classifier to perform scale selection in the linear and morphological scale spaces. Thereby, the best (implicit) trade-off

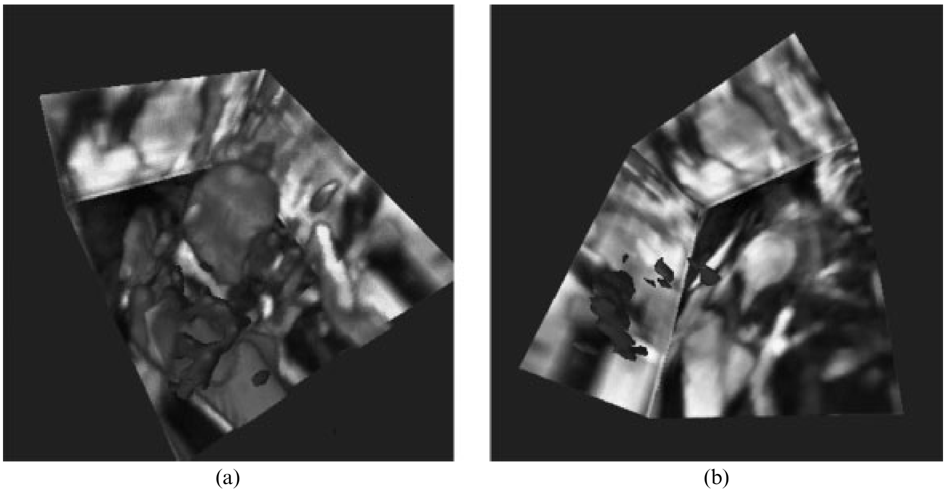


Figure 5. Visualizations of a 3D MRI dataset showing nests with viable tumour in a femur (hip): (a) the nests shown with dark-grey have been detected by a neural network. The light-grey (iso-)surfaces indicate the healthy bone marrow inside the femur. The sides and the bottom show maximal intensity projections of the 3D MRI dataset; (b) the bone marrow is omitted, allowing the observer to obtain a 3D impression of the size of the remnants with viable tumour.

between location bias and variance is being made for a given test set.

Lindeberg showed that it is difficult to identify the optimal scale for a particular feature detector.[33] It should be kept in mind that the optimal set of scales depends on the amount of noise present in the images and the typical size of the (connected) segments one wants to find. In general, finding an *analytical* solution to this problem may be intractable. Our approach to scale selection is a pragmatic alternative, which may be applied to signal-driven segmentation algorithms based on spatial information from the linear or morphological scale space. The approach results in a particular trade-off between (location) bias and variance, namely that which minimizes the error rate on a representative test set. One may prefer a smooth, labelled image with a high local confidence and a smaller correctness over a labelled image with more scatter and a higher number of correctly segmented voxels. The optimal trade-off between such conflicting criteria can only be determined by the end user. If one criterion is being used, the optimal sampling of the scale space can be found through experiments.

We introduced isotropic quality measures for studying the effect of varying the scale of the images provided as input to the neural net classifier. Such measures make it possible to quantify homogeneity aspects of a segmentation result, which may otherwise have been left for subjective assessment; for a discussion see, for example, Zhang.[49] We feel that especially the local confidence measure results in plausible measures for the amount of scatter in a labelled image.

It is clear that for a signal-based segmentation classifier to be uncommitted with respect to scale, the prior class distribution in the training set needs to be uniform. This is a direct consequence of Theorem 1, Lemma 1 and Corollary 1. Experiments reported elsewhere indicate that, in general, for the linear discriminant and classification based on logistic regression,[50] the best training result is obtained for a uniform prior distribution. The experiment with the synthetic image data confirms that the class-conditional error rates remain constant for a varying prior. Consequently, we recommend not using the formula of McMichael, equation (13), to correct for a different prior distribution, unless there is a clear expectation regarding the prior probability distribution of the different types of segments in a particular image. In other words, we advocate for an uncommitted segmentation approach based on a uniform prior class distribution. In a particular application, it may be desirable to optimize the performance, thereby choosing a representative (non-uniform) class distribution.

It is well known that trained classifiers suffer from the curse of dimensionality, which impedes generalization when the number of features becomes high. This so-called peaking phenomenon[51,52] implies an increasing difficulty in discerning discriminative from useless features as the dimensionality of the feature space increases.[53] The peaking phenomenon can prevent our scale selection algorithm from choosing the best set of scales.

Conclusion

In this article, we have analysed the problem of scale selection for signal-driven segmentation algorithms based on pattern classifiers. Theoretical results indicate that, in the presence of noise, the sampling of the (discrete) linear scale space entails a trade-off between (location) bias and variance. Based on this analysis, we propose to use the overall error rate obtained on a test set to optimize the sampling of the scale space. It is furthermore shown that the class-conditional error rate (per type of segment) remains constant per unit of area under zooming. This advocates for building an uncommitted signal-driven segmentation approach based on a uniform prior class distribution in the training set.

The optimal set of scales depends on several factors including the noise level present in the image material, the prior distribution of the different types of segments, the class-conditional distributions associated with each type of segment as well as the actual size of the (connected) segments. Often, conflicting criteria need to be fulfilled in order to obtain the best possible trade-off between variance and location bias. Experiments with a neural net classifier developed for segmentation of dynamic MR images illustrate these results. The experiments also show that adding spatial features to the classifier, extracted from the linear or morphological scale spaces, improves the segmentation result compared to a signal-driven approach based solely on the dynamic MR signal. The performance on a set of test images is used to select the two scales that result in the best performance.

Two novel spatial quality measures were introduced, both characterizing spatial properties of a labelled image. These measures as well as the known statistical quality measures correctness and kappa, have been used to quantify the improvement of the obtained segmentation result. According to the computed quality measures, the linear scale space is the best configuration for this tumour tissue classification task.

ACKNOWLEDGEMENTS

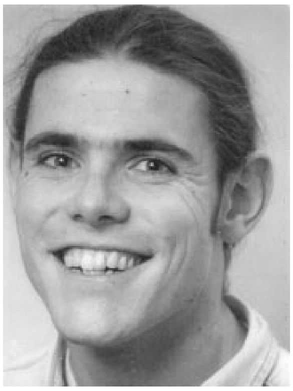
We are grateful to Prof. Dr. J. L. Bloem (Department of Radiology) from the Leiden University Medical Centre, for providing us with the patient material. This project was financially supported by the Dutch Cancer Foundation (KWF), Grant RUL 97–1509.

References

1. Fukumi M, Omatu S, Takeda F, *et al*. Rotation-invariant neural pattern-recognition system with application to coin recognition. *IEEE Transactions on Neural Networks* 1992; **3**(2): 272–279.
2. DeKruger D, Hunt BR. Image processing and neural networks for recognition of cartographic area features. *Pattern Recognition* 1994; **27**(4): 461–483.
3. Ozkan M, Dawant BM, Maciunas RJ. Neural-network-based segmentation of multi-modal medical images: a comparative and prospective study. *IEEE Transactions on Medical Imaging* 1993; **12**(3): 534–544.
4. Kung SY, Taur JS. Decision based neural networks with signal image classification applications. *IEEE Transactions on Neural Networks* 1995; **6**(1): 170–181.
5. Laine A, Fan J. Texture classification by wavelet packet signatures. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1993; **15**(11): 1186–1191.
6. Reddick WE, Glass JO, Cook EN, *et al*. Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks. *IEEE Transactions on Medical Imaging* 1997; **16**(6): 911–918.
7. Egmont-Petersen M, Arts T. Recognition of radiopaque markers in X-ray images using a neural network as nonlinear filter. *Pattern Recognition Letters* 1999; **20**(5): 521–533.
8. Egmont-Petersen M, Pelikan E. Detection of bone tumours in radiographs using neural networks. *Pattern Analysis and Applications* 1999; **2**(2): 172–183.
9. Frangi AF, Egmont-Petersen M, Niessen WJ, *et al*. Bone tumor segmentation in MR perfusion images with neural networks using multiscale pharmacokinetic features. *Image and Vision Computing* 2001; **19**(9–10): 679–690.
10. Fukumi M, Omatu S, Nishikawa Y. Rotation-invariant neural pattern recognition system estimating a rotation angle. *IEEE Transactions on Neural Networks* 1997; **8**(3): 568–581.
11. Hu MK. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory* 1962; **8**(2): 179–187.
12. Zernike F. Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form der Phasenkontrastmethode. *Physica* 1934; **1**: 689–701.
13. Persoon E, Fu K-S. Shape discrimination using Fourier descriptors. *IEEE Transactions on Systems, Man and Cybernetics* 1977; **7**(3): 170–179.
14. Lehmann TM, Gonner C, Spitzer K. Survey: interpolation methods in medical image processing. *IEEE Transactions on Medical Imaging* 1999; **18**(11): 1049–1075.
15. Ridder DD, Kittler J, Lemmers O, *et al*. The adaptive subspace map for texture segmentation. In *Proceedings of the 15th International Conference of Pattern Recognition*, Barcelona, 2000; 216–220.
16. Chen Q, Defrise M, Deconinck F. Symmetric phase-only matched filtering of Fourier–Mellin transforms for image registration and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1994; **16**(12): 1156–1168.
17. Götze N, Drüe S, Hartmann G. Invariant object recognition with discriminant features based on local Fast Fourier Mellin Transform. In *Proceedings of the 15th International Conference of Pattern Recognition*, Barcelona, 2000; 948–951.
18. Romeny BMT, Florack LMJ, Salden AH, *et al*. Higher-order differential structure of images. *Image and Vision Computing* 1994; **12**(6): 317–325.
19. Jiang XY, Bunke H. Simple and fast computation of moments. *Pattern Recognition* 1991; **24**(8): 801–806.
20. Liao SX, Pawlak M. On image analysis by moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1996; **18**(3): 254–266.
21. Egmont-Petersen M, Frangi AF, Niessen WJ, *et al*. Segmentation of bone tumor in MR perfusion images using neural networks and multiscale pharmacokinetic features. In *Proceedings of the 15th International Congress on Pattern Recognition*, Barcelona, 2000; 80–83.
22. Duda RO, Hart PE. *Pattern Classification and Scene Analysis*. Wiley: New York; 1973.
23. Egmont-Petersen M, Talmon JL, Hasman A, *et al*. Assessing the importance of features for multi-layer receptrons. *Neural Networks* 1998; **11**(4): 623–635.
24. Florack LMJ, ter Haar Romeny BM, Koenderink JJ, *et al*. Scale and the differential structure of images. *Image and Vision Computing* 1992; **10**(6): 376–388.
25. Koenderink JJ. The structure of images. *Biological Cybernetics* 1984; **50**: 363–370.
26. Lindeberg T. *Scale-Space Theory in Computer Vision*. Kluwer: Boston, MA, 1994.
27. Witkin A. Scale-dependent qualitative signal description. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, Karlsruhe, 1983; 1019–1022.
28. Schumaker LL. *Spline Functions: Basic Theory*. Wiley: New York, 1981.
29. Ralston A. *A First Course in Numerical Analysis*. McGraw-Hill: Tokyo, 1965.
30. McMichael DW. BARTIN: minimizing Bayes risk and incorporating priors using supervised learning networks. *IEE Proceedings—F Radar and Signal Processing* 1992; **139**(6): 413–419.
31. Lindeberg T. Effective scale: a natural unit for measuring scale-space lifetime. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1993; **15**(10): 1068–1074.
32. Florack LMJ. The syntactical structure of scalar images. Thesis, Image Sciences Institute, Utrecht University, Utrecht, 1993.

33. Lindeberg T. Feature detection with automatic scale selection. *International Journal of Computer Vision* 1998; **30**(2): 79–116.
34. Egmont-Petersen M, Talmon JL, Hasman A. Robustness metrics for measuring the influence of additive noise on the performance of statistical classifiers. *International Journal of Medical Informatics* 1997; **46**(2): 103–112.
35. Egmont-Petersen M, Talmon JL, Brender J, *et al*. On the quality of neural net classifiers. *Artificial Intelligence in Medicine* 1994; **6**(5): 359–381.
36. Egmont-Petersen M, Pelikan E. Erweiterte Kriterien zur Beurteilung von Segmentierungsergebnissen. In *Proceedings of the 4th Workshop on Digital Image Processing in Medicine*, Freiburg, 1996; 24–30.
37. Richard MD, Lippmann RP. Neural network classifiers estimate bayesian posterior probabilities. *Neural Computation* 1991; **3**(4): 461–483.
38. Funahashi K-I. Multilayer neural networks and Bayes decision theory. *Neural Networks* 1998; **11**(2): 209–213.
39. Funahashi K-I. On the approximate realization of continuous mappings by neural networks. *Neural Networks* 1989; **2**(3): 183–192.
40. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Networks* 1989; **2**(5): 359–366.
41. Chong C, Jia J. Assessments of neural network classifier output codings using variability of Hamming distance. *Pattern Recognition Letters* 1996; **17**(8): 811–818.
42. Egmont-Petersen M, Hogendoorn PCW, Geest R van der, *et al*. Detection of areas with viable remnant tumor in postchemotherapy patients with Ewing's sarcoma by dynamic contrast-enhanced MRI using pharmacokinetic modeling. *Magnetic Resonance Imaging* 2000; **15**(5): 525–535.
43. Verbeek PW, Vrooman HA, Vliet LJ van. Low-level image-processing by max min filters. *Signal Processing* 1988; **15**(3): 249–258.
44. Alpert NM, Bradshaw JF, Kennedy D, *et al*. The principle axes transformation: a method for image registration. *Journal of Nuclear Medicine* 1990; **31**(10): 1717–1722.
45. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In *Explorations in the microstructure of Cognition*. Vol. I: *Parallel Distributed Processing*, Rumelhart DE, McClelland JL (eds). MIT Press: Cambridge, MA, 1986; 319–362.
46. Jackway PT, Deriche M. Scale-space properties of the multiscale morphological dilation–Erosion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1996; **18**: 38–51.
47. Park K-R, Lee C-N. Scale-space using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1996; **18**: 1121–1126.
48. Karthaus V, Thygesen H, Egmont-Petersen M, *et al*. User-requirements driven learning. *Computer Methods and Programs in Biomedicine* 1995; **48**(1–2): 39–44.
49. Zhang YJ. A survey on evaluation methods for image segmentation. *Pattern Recognition* 1996; **29**(8): 1335–1346.
50. Kao TC, McCabe GP. Optimal sample allocation for normal discrimination and logistics-regression under stratified sampling. *Journal of the American Statistical Association* 1991; **86**(414): 432–436.
51. Chandrasekaran B, Jain AK. Independence, measurement complexity, and classification performance. *IEEE Transactions on Systems, Man and Cybernetics* 1975; **5**(2): 240–244.
52. Waller WG, Jain AK. On the monotonicity of the performance of a Bayesian classifier. *IEEE Transactions on Information Theory* 1978; **24**(3): 392–394.
53. Hamamoto Y, Uchimura S, Tomita S. On the behavior of artificial neural network classifiers in high-dimensional spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1996; **18**(5): 571–574.

Authors' biographies:



J. P. Janssen was born in Heteren, the Netherlands, in 1976. He received his MSc degree in electrical engineering from the Delft University of Technology in 1999. He is currently associated with the Division of Image Processing, Department of Radiology, Leiden University Medical Centre, as a PhD student. He is currently working on the border detection of coronary arteries in X-ray angiograms. His main research interests are image segmentation and image classification.

Dr Egmont-Petersen was born in Copenhagen, Denmark, in 1967. He received the combined BS and MS degrees in computer science/business administration from Copenhagen Business School in 1988 and 1990, respectively. He received the PhD degree in medical informatics from Maastricht University, the Netherlands, in 1996. He has worked from 1997 to 2000 for the Division of Image Processing, Department of Radiology, Leiden University Medical Centre, as a postdoctoral researcher. He is currently associated with the Department of Computer Science at the University of Utrecht, the Netherlands. Dr Egmont-Petersen is currently developing novel learning algorithms for Bayesian belief networks. His main research interests include belief networks, neural networks, support vector machines, statistical classifiers, feature selection, image understanding and invariant theory. He has published more than 35 papers in journals and conference proceedings.

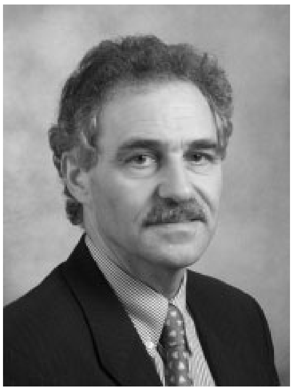
E. A. Hendriks received his MSc and PhD degrees from the University of Utrecht in 1983 and 1987, respectively, both in physics. In 1987 he joined the Electrical Engineering Faculty of Delft University of Technology as an assistant professor. In 1994 he became a member of the Information and Communication Theory of this faculty and since 1997 he has headed the computer vision section of this group as an associate professor. His interest is in low-level image processing, image segmentation, stereoscopic and 3D imaging, motion and disparity estimation and real-time applications.

M. J. T. Reinders is an associate professor at the Delft University of Technology in the Information and Communication Theory Group of the Electrical Engineering Department. He is active on the bridge between the fields of computer vision and that of machine intelligence. These activities give him a specific expertise on the use model and/or knowledge information in computer vision, supervised and unsupervised pattern recognition, as well as machine learning and data-mining techniques. Recently, he became interested in the rapidly growing field of bioinformatics, which uses machine-learning techniques to study biological phenomena such as the behaviour of gene expressions and protein mapping.

R. J. van der Geest received the MSc degree from Delft University of Technology, Department of Electrical

Engineering. Since 1992, he has been working as a senior research associate at the Division of Image Processing, supervising the development of segmentation algorithms for the analysis of cardiovascular magnetic resonance images. His main expertise lies in the field of cardiovascular MR image analysis, model-based segmentation and 3D visualization.

P. C. W. Hogendoorn was born in Leiden, the Netherlands, in 1960. He received his MSc degree followed by his MD degree from Leiden University in 1986 and 1989 respectively. In 1990 he received his PhD degree in experimental immunology. Following a clinicopathological fellowship at the Dutch National Cancer Institute, Amsterdam (1992) and the Memorial Sloan Kettering Cancer Centre, New York (1993), he was trained during his residency at the Leiden University Medical Centre (LUMC) as a pathologist. In 1994 he was appointed on a tenured position as staff pathologist at LUMC. His research and clinicopathological focus turned to (molecular) pathology and imaging of bone and soft tissue tumours. In 1995 he was appointed as a member of the Netherlands Committee of Bone Tumours, followed by an appointment as chairman of the pathology and biology subcommittee of the Bone and Soft Tissue Tumour Study Group of the EORTC and Osteosarcoma Intergroup in 1997. Since 1998 he has been a full professor of pathology at Leiden University.



J. H. C. Reiber was born in Haarlem, the Netherlands, in 1946. He received his MScEE degree from the Delft University of Technology in 1971 and the PhD degree in electrical engineering in 1975 from Stanford University, California, USA. In 1977 he founded the Laboratory for Clinical and Experimental Image Processing (LKEB) at the Thorax Centre in Rotterdam, directing the research at the development and validation of objective and automated techniques for the segmentation of cardiovascular images, in particular for quantitative coronary arteriography (QCA), nuclear cardiology and echocardiology. With the move of LKEB in 1990 to the Leiden University Medical Centre (LUMC), the scope broadened to intravascular ultrasound, MRI, CT, etc., also in radiological applications. Since April 1995 he has been a professor of medical image processing, in particular in cardiovascular applications, at the LUMC and the Interuniversity Cardiology Institute of the Netherlands. His research interests include (knowledge-guided) image processing and its clinical applications.