Segmentation of Ultrasonic Images Using Support Vector Machines [*]

Constantine Kotropoulos [a],[*] Ioannis Pitas [a]

[a] *Department of Informatics, Box 451 Aristotle University of Thessaloniki, Thessaloniki 540 06, Greece.*

Abstract

Support Vector Machines are a general algorithm based on guaranteed risk bounds of statistical learning theory. They have found numerous applications, such as in classification of brain PET images, optical character recognition, object detection, face verification, text categorization and so on. In this paper we propose the use of support vector machines to segment lesions in ultrasound images and we assess thoroughly their lesion detection ability. We demonstrate that trained support vector machines with a Radial Basis Function kernel segment satisfactorily (unseen) ultrasound B-mode images as well as clinical ultrasonic images.

Key words: ultrasound B-mode images, lesion detection, segmentation, Support Vector Machines, Radial Basis Function kernel

1 Introduction

Support Vector Machines (SVMs) are a state-of-the-art pattern recognition technique whose foundations stem from statistical learning theory (Vapnik, 1995, 1998; Cristianini and Shawe-Taylor, 2000). However, the scope of SVMs goes beyond pattern recognition, because they can also handle two more learning problems, i.e., regression estimation and density estimation. An SVM is a general algorithm based on guaranteed risk bounds of statistical learning

theory, i.e., the so-called *structural risk minimization principle*. It is a learning machine capable of implementing a set of functions that approximate best the supervisor's response with an expected risk bounded by the sum of the empirical risk and the Vapnik-Chervonenkis (VC) confidence. The latter is a bound on the generalization ability of the learning machine, that depends on the so-called VC dimension of the set of functions implemented by the machine. There is a close relationship between SVMs and regularization networks (Evgeniou et al., 2000b) and the Radial Basis Function (RBF) classifiers (Schoelkopf et al., 1997). Among the many tutorials on SVMs that can be found in literature, we refer to (Burges, 1998; Osuna et al., 1997). SVMs have found numerous applications such as in optical character recognition, object detection, face verification, text categorization, and so on. In the field of medical imaging they have been applied to the classification of brain PET volumes (Bonneville et al., 1998). Another relevant application of SVMs is in breast cancer diagnosis and prognosis where the Wisconsin breast cancer dataset was used for a binary classification task (the tumor is malignant or benign) (Mangasarian et al., 1995).

Two approaches in the segmentation of ultrasound B-mode images can be identified in the related literature. The first approach tries to reduce the speckle so that the success of any segmentation algorithm applied to the filtered image afterward increases. In this approach detection (i.e., segmentation) and estimation (i.e., filtering) are treated in parallel (Kotropoulos and Pitas, 1992; Kotropoulos et al., 1994). Another approach convolves the amplitudes of the ultrasound B-mode image with a block shaped kernel whose domain and height are modulated by the local signal level (Collaris and Hoeks, 1995). Adaptive smoothing techniques could also be applied (Karaman et al., 1995). Log-compression is shown to enhance the weak backscatterers (Dutt and Greenleaf, 1996). The authors have derived the statistics of log-compressed echo images using the K-distribution model for the echo envelope in order to quantify the extent of speckle formation and to reduce it adaptively. An analog cellular neural network algorithm that combines optimal nonlinear filtering, fuzzy logic, and constrained wave propagation in order to estimate the continuous contour of the inner boundary of the left ventricle from a sequence of echocardiographic images is described in (Rekeczky et al., 1999). The second approach deals only with the segmentation problem taking into account the speckle properties. The segmentation of echographic images using mathematical morphology was studied in (Klingler et al., 1988). A multiresolution texture segmentation approach that addresses texture characterization, image resolution, and the time to complete the segmentation was developed in (Muzzolini et al., 1993). Maximum a posteriori probabilistic segmentation of ultrasound images was treated in (Ashton and Parker, 1995). Methods to evaluate the performance of segmentation algorithms in synthetic aperture radar images were developed based on the properties of coherent speckle in (Caves et al., 1998).

2

In this paper, we propose the application of SVMs to the segmentation of ultrasonic images. SVMs have successfully been applied to object detection tasks (Papageorgiou and Poggio, 1999; Evgeniou et al., 2000a; Mohan et al., 2001). In these tasks a rich, overcomplete dictionary of Haar wavelet features was used to describe an object class. Special effort has been paid toward the thorough and compact representation of the object class that encompasses the geometric constraints for each class (centroid, top and bottom boundary edges, permissible scales). This is not the case with our work. We have not employed any wavelet features and have not made any attempt to derive any representation of the lesion area, because the lesion areas vary in different ultrasound images. We are interested in learning the local signal properties that discriminate the lesion areas from the background in tissue mimicking ultrasound B-mode images and to exploit these properties to segment clinical ultrasonic images. Accordingly, we assess the performance of SVMs on ultrasound B-mode images that simulate a homogeneous piece of tissue with a circular lesion in the middle (Verhoeven et al., 1991) using several window sizes and compare the performance of the proposed method to that of thresholding the L_2 mean filtered image (Kotropoulos and Pitas, 1992) and the method proposed by Collaris and Hoeks (Collaris and Hoeks, 1995) using the area under the Receiver Operating Characteristic (ROC) and the probability of detection for a fixed probability of false alarm as figures of merit. We consider both light and dark lesions as well as cases where the lesion has an equal number density of scatterers to the background or a lower number density of scatterers than the background. We demonstrate that SVMs with Radial Basis Function kernels (RBF-SVMs) segment the ultrasound B-mode images more effectively than thresholding the L_2 mean filtered image. Their generalization ability is assessed by applying the trained RBF-SVMs on test images different from those used in the training phase.

The outline of the paper is as follows. The foundations of support vector machines are described in Section 2. Experimental results are presented in Section 3 and conclusions are drawn in Section 4.

2   Support Vector Machines

The segmentation of simulated ultrasound B-mode images into lesion and background regions can easily be described as a binary detection problem. Let N_1 and N_2 denote the image rows and columns, respectively. By moving a running window of dimensions $m \times m$ over the image we collect the $M = m^2$ gray level values of the pixels that fall within the window and we create an input vector $\mathbf{x}_i \in I\!\!R^M$, where $i = (\iota_1 - 1)N_2 + \iota_2$ is a running index that corresponds to the coordinates (ι_1, ι_2) of the central pixel of the running window for $1 \leq \iota_1 \leq N_1$ and $1 \leq \iota_2 \leq N_2$. Image segmentation can be

formulated as the estimation of a function $f : \mathbb{R}^M \to \{-1, +1\}$ using input-output independent identically distributed (i.i.d) training data pairs generated according to an unknown probability distribution $P(\mathbf{x}, y)$

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_L, y_L) \in \mathbb{R}^M \times \{-1, 1\} \tag{1}$$

such that f correctly classifies the unseen test data $(\mathbf{x}, y)$. An input vector $\mathbf{x}$ is assigned to the class $y = +1$ if $f(\mathbf{x}) > 0$ and to the class $y = -1$ otherwise. In the context of ultrasound image segmentation we agree that the class $y = +1$ refers to the lesion, while the class $y = -1$ refers to the background. We further assume that the test data are generated by the same probability distribution $P(\mathbf{x}, y)$ as the training examples. The best segmentation function f is chosen so that the expected risk

$$R[f] = \int \ell(f(\mathbf{x}, y)) dP(\mathbf{x}, y) \tag{2}$$

is minimized where $\ell(\cdot)$ denotes an appropriate loss function, e.g. the zero-one loss $\ell(f(\mathbf{x}, y)) = u(-yf(\mathbf{x}))$ with $u(t)$ representing the unit-step function. Since the underlying probability distribution $P(\mathbf{x}, y)$ is not known, the expected error (2) cannot be minimized directly. We have to estimate a function that is close to the optimal one based on the training data pairs and the properties of the class of functions where f belongs to. The approximation of the expected risk (2) by the *empirical risk*

$$R_{\text{emp}}[f] = \frac{1}{L} \sum_{i=1}^{L} \ell(f(\mathbf{x}_i, y_i)) \tag{3}$$

and the attempt to find the function f that minimizes (3) does not solve the problem, because the empirical risk converges toward the expected risk only if $L \to \infty$. That is, for small sample sizes there might be large deviations between the expected and the empirical risk. Overfitting might also occur yielding a large generalization error, despite the fact that f commits the smallest number of errors in the training set. To avoid overfitting, one should restrict the complexity of the class of functions where f comes from (Vapnik, 1995, 1998). Vapnik-Chervonenkis (VC) theory controls the complexity of a function class through the notion of the VC dimension h of the class of functions the estimate f is chosen from and aims at finding the function f that minimizes the upper bound on the expected risk given by (Vapnik, 1995)

$$R[f] \le R_{\text{emp}}[f] + \sqrt{\frac{h(\ln(\frac{2L}{h}) + 1) - \ln(\frac{\eta}{4})}{L}} \tag{4}$$

4

that holds with probability $1 - \eta$ for $0 \leq \eta \leq 1$. The second term on the right hand side of (4) is the so called VC confidence. There are two strategies for minimizing the upper bound. The first one is to keep the VC confidence fixed and to minimize the empirical risk and the second one is to fix the empirical risk to a small value and to minimize the VC confidence. The latter approach is the so-called *structural risk minimization* principle. It formally implies that a structure should a priori be defined in the space of the class of functions, before the training data appear. Although SVMs implement this principle, their training algorithm that aims to minimize the VC dimension is based on an hierarchy that depends on the data (Vapnik, 1995). Fat shattering bounds on the generalization error, $R[f] - R_{\mathrm{emp}}[f]$, are another type of bounds that also depend on the data (Shawe-Taylor et al., 1998; Cristianini and Shawe-Taylor, 2000). From a practical point of view, leave-one-out cross-validation estimates of the empirical risk are more useful, because they can be exploited to optimally choose the model parameters (Vapnik, 1998; Lee and Lin, 2001). Frequently, all the aforementioned bounds are loose. In the following, we elaborate on the training algorithm of SVMs. The starting point is the construction of the optimal hyperplane for linearly separable training vectors $\mathbf{x}_i$. The extension to nonseparable training sets as well as to the nonlinear case is treated next.

2.1 Separable case

If the training set is separable by a hyperplane, then there exists a function of the form

$$f_{\mathbf{w},b}(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = 0 \tag{5}$$

that satisfies the following inequalities:

$$\begin{aligned} f_{\mathbf{w},b}(\mathbf{x}_i) &\geq 1 \quad \text{if } y_i = 1 \\ f_{\mathbf{w},b}(\mathbf{x}_i) &\leq -1 \quad \text{if } y_i = -1 \end{aligned} \tag{6}$$

where $\mathbf{w}$ is the vector of hyperplane coefficients and b is a bias term. In compact notation, the set of inequalities (6) can be rewritten as

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 \geq 0 \quad i = 1, 2, \ldots, L. \tag{7}$$

The optimal separating hyperplane is the one that maximizes the margin between itself and the separated vectors (Vapnik, 1998). The margin is given

by

$$\rho(\mathbf{w}) = \frac{1}{2}\left(\min_{i:y_i=1}(\frac{f_{\mathbf{w},b}(\mathbf{x}_i)}{\|\mathbf{w}\|}) - \max_{i:y_i=-1}(\frac{f_{\mathbf{w},b}(\mathbf{x}_i)}{\|\mathbf{w}\|})\right) = \frac{1}{\|\mathbf{w}\|}. \tag{8}$$

Accordingly, the optimal separating hyperplane is the one that minimizes the norm

$$\|\mathbf{w}\|^2 = \mathbf{w}^T\mathbf{w} \tag{9}$$

subject to the inequalities (7). From an algorithmic point of view, the minimization of the objective function (9) subject to the constraint functions (7), is a constrained quadratic optimization problem (Fletcher, 1987). The solution of the optimization problem under study is given by the saddle point of the Lagrangian

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \mathbf{w}^T\mathbf{w} - \sum_{i=1}^{L} \alpha_i \left(y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \right) \tag{10}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_L)^T$ is the vector of Lagrange multipliers. The Lagrangian has to be minimized with respect to $\mathbf{w}$ and b and maximized with respect to $\alpha_i > 0$. The Kuhn-Tucker (KT) conditions imply that (Fletcher, 1987)

$$\nabla_{\mathbf{w}} L(\mathbf{w}_o, b_o, \boldsymbol{\alpha}_o) = \mathbf{0} \Leftrightarrow \mathbf{w}_o = \frac{1}{2} \sum_{i=1}^{L} \alpha_{i,o} y_i \mathbf{x}_i$$

$$\frac{\partial}{\partial b} L(\mathbf{w}_o, b_o, \boldsymbol{\alpha}_o) = 0 \Leftrightarrow \sum_{i=1}^{L} \alpha_{i,o} y_i = 0$$

$$y_i \left(\mathbf{w}_o^T \mathbf{x}_i + b_o \right) - 1 \geq 0 \qquad i = 1, 2, \ldots, L \tag{11}$$

$$\alpha_{i,o} \geq 0 \qquad i = 1, 2, \ldots, L$$

$$\alpha_{i,o} \left(y_i \left(\mathbf{w}_o^T \mathbf{x}_i + b_o \right) - 1 \right) = 0 \qquad i = 1, 2, \ldots, L.$$

From the conditions (11), one can see that the vector of hyperplane coefficients we search for, is a linear combination of the training vectors. Moreover, it is the linear combination of the training vectors whose associated Lagrange multipliers are nonzero. These training vectors are called *support vectors* (SVs) (Vapnik, 1995). Putting the expression for $\mathbf{w}_o$ into the Lagrangian (10) and taking into account the KT conditions, we obtain the Wolfe dual functional

$$\mathcal{W}(\boldsymbol{\alpha}) = \sum_{i=1}^{L} \alpha_i - \frac{1}{4} \sum_{i=1}^{L} \sum_{j=1}^{L} \alpha_i \alpha_j \underbrace{y_i y_j \left(\mathbf{x}_i^T \mathbf{x}_j \right)}_{H_{ij}} \tag{12}$$

where H_{ij} is the ijth element of the Hessian matrix $\mathbf{H}$. The maximization of (12) in the non-negative quadrant of α_i, i.e.

$$\alpha_i \geq 0 \qquad i = 1, 2, \ldots, L \tag{13}$$

under the constraint

$$\sum_{i=1}^{L} \alpha_i y_i = 0 \tag{14}$$

is equivalent to the optimization problem

$$\text{minimize} \quad \frac{1}{4} \boldsymbol{\alpha}^T \mathbf{H} \, \boldsymbol{\alpha} - \mathbf{1}^T \boldsymbol{\alpha} \ \text{ subject to (13) and (14)} \tag{15}$$

where $\mathbf{1}$ denotes a vector of ones. The optimization problem (15) can be solved by using any optimization software package, e.g. (Joachims, 1998). For a review of the optimization algorithms, the interested reader is referred to (Osuna et al., 1997). Having found the nonzero Lagrange multipliers, $\alpha_{i,o}$, the optimal separating hyperplane is given by

$$f(\mathbf{x}) = \text{sgn}\left(\underbrace{\frac{1}{2} \sum_{\alpha_{i,o} > 0} y_i \, \alpha_{i,o} \, \mathbf{x}_i^T}_{\mathbf{w}_o^T} \mathbf{x} + b_o \right) \tag{16}$$

where sgn() denotes the sign function. The bias term, b_o, can be obtained from a pair of support vectors $(\mathbf{x}_+, \mathbf{x}_-)$ that belong to different classes as follows:

$$b_o = -\frac{1}{2} \sum_{j=1}^{L} \alpha_{j,o} \, y_j \, \mathbf{x}_j^T \, (\mathbf{x}_+ + \mathbf{x}_-). \tag{17}$$

2.2 Linearly Nonseparable Case

When the training vectors are not linearly separable, we would like to relax the constraints (6) by introducing nonnegative slack variables, ξ_i, $i = 1, 2, \ldots, L$, to allow the possibility of examples that violate (7) (Vapnik, 1995)

$$\begin{aligned} \mathbf{w}^T \mathbf{x}_i \geq 1 - b - \xi_i \qquad & \text{if } y_i = 1 \\ \mathbf{w}^T \mathbf{x}_i \leq b - 1 + \xi_i \qquad & \text{if } y_i = -1. \end{aligned} \tag{18}$$

The so-called generalized optimal hyperplane is determined by the vector $\mathbf{w}_o$ that minimizes the functional

$$J(\mathbf{w}, b, \boldsymbol{\xi}) = \mathbf{w}^T \mathbf{w} + C (\sum_{i=1}^{L} \xi_i)^\sigma, \qquad \sigma > 0 \tag{19}$$

subject to:

$$\xi_i \geq 0 \qquad i = 1, 2, \ldots, L \tag{20}$$

where C is a parameter chosen by the user that defines the cost of constraint violations. The larger the parameter C is, the higher penalty to the errors is assigned. The minimization of (19) subject to (18) and (20) is a convex programming problem for any integer σ. For $\sigma = 1$ or 2, it is a quadratic programming problem. Moreover, the choice $\sigma = 1$ has the advantage that neither ξ_t nor their Lagrange multipliers appear in the Wolfe dual problem (Burges, 1998). The Lagrangian of the optimization problem is given by:

$$L(\mathbf{w}, b, \boldsymbol{\alpha}, \boldsymbol{\xi}, \boldsymbol{\mu}) = \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{L} \xi_i - \sum_{i=1}^{L} \alpha_i \left\{ y_i(\mathbf{w}^T \mathbf{x}_i + b) + \xi_i - 1 \right\}$$
$$- \sum_{i=1}^{N} \mu_i \xi_i \tag{21}$$

where μ_i are Lagrange multipliers enforcing the positivity of ξ_i. To find the coefficients of the generalized optimal hyperplane that satisfy $\nabla_{\mathbf{w}} L(\mathbf{w}_o, b_o, \boldsymbol{\alpha}_o, \boldsymbol{\xi}_o, \boldsymbol{\mu}_o) = \mathbf{0}$, one has to find the Lagrange multipliers, α_i, $i = 1, 2, \ldots, L$, that maximize the Wolfe dual problem

$$\mathcal{W}(\boldsymbol{\alpha}) = \mathbf{1}^T \boldsymbol{\alpha} - \frac{1}{4} \boldsymbol{\alpha}^T \mathbf{H} \boldsymbol{\alpha} \text{ subject to } 0 \leq \alpha_i \leq C \ i = 1, 2, \ldots, L$$
$$\text{and } \sum_{i=1}^{L} \alpha_i y_i = 0. \tag{22}$$

The comparison of (22) and (15) reveals that the objective function and the equality constraint (14) remain unchanged, while the Lagrange multipliers are now upper bounded by C. As in the separable case, only some of the Lagrange multipliers α_i are nonzero. These multipliers are used to determine the support vectors. There are either support vectors on the margin that satisfy $0 < \alpha_i < C$ and $y_i f(\mathbf{x}_i) = 1$ or inside the margin area that satisfy $\alpha_i = C$ and $y_i f(\mathbf{x}_i) < 1$ (Müller et al., 2001). Having determined the support vectors, $\mathbf{w}_o$ is derived by the first equation in (11). A numerically stable solution for the bias term

is given by (Joachims, 1998; Müller et al., 2001)

$$b = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \left(y_i - \sum_{j=1}^{L} y_j \alpha_j \mathbf{x}_i^T \mathbf{x}_j \right) \qquad (23)$$

where $\mathcal{I} = \{i : 0 < \alpha_i < C\}$ and $|\mathcal{I}|$ denotes the cardinality of the set $\mathcal{I}$.

2.3 Nonlinear Support Vector Machines

Thus far, we have described the case of hyperplanes (i.e., linear decision surfaces). By examining (22), one may notice that inner products of the training vectors appear in the Wolfe dual problem, the expression for the optimal hyperplane coefficient vector $\mathbf{w}_o$, and the bias term b. To allow for a more complex decision surface, the training vectors $\mathbf{x}_i$, $i = 1, 2, \ldots, L$ are nonlinearly mapped onto a high-dimensional feature space by $\Phi : I\!\!R^M \mapsto \mathcal{F}$ and then a linear separation is attempted in $\mathcal{F}$. If $\mathcal{F}$ is a Hilbert space, thanks to Hilbert-Schmidt theory there exists a kernel function in the original space $I\!\!R^M$ that describes the inner product in $\mathcal{F}$ (Vapnik, 1995, 1998), i.e.

$$< \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) >= K(\mathbf{x}_i, \mathbf{x}_j) \qquad (24)$$

where $K(\mathbf{x}_i, \mathbf{x}_j)$ should satisfy Mercer's condition. That is, for any kernel function satisfying Mercer's condition, there exists a space $\mathcal{F}$ where this function generates the inner product. Functions that are usually employed in constructing nonlinear SVMs are the polynomial kernel, the RBF kernel, the sigmoid kernel, etc. The kernel function mostly used in the experiments reported in Section 3 is the RBF one defined formally as

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2). \qquad (25)$$

3 Experimental results

Simulated ultrasound B-mode images are used in order to evaluate the performance of various image processing algorithms and to select the parameters involved in them. Such images are suitable to train an algorithm, like the SVM, that is based on examples and counterexamples, because there we know where the lesion and the background are. We trained SVMs using the RBF kernel defined in (25) on training sets selected from simulated ultrasound B-mode images and we tested their performance on test sets that include unseen input vectors from either the same image or a different one. In the training phase we

used the set of simulated ultrasound B-mode images described in (Verhoeven et al., 1991). They are simulations of a homogeneous piece of tissue $(4 \times 4 \text{ cm}^2)$ with a lesion in the middle having a diameter of 2 cm. The lesion differs from the background in the reflection strength, the number of scatterers, or both. The background has a number density of scatterers 5000 cm^{-3} that yields a fully developed speckle. The lesion has a number density of scatterers of either 500 cm^{-3} that yields a sub-Rayleigh distribution or 5000 cm^{-3} that yields pixel gray values that are Rayleigh distributed. In the latter case, the second order statistics are the same in the lesion and the background areas. Table 1 summarizes the characteristics of the simulated B-mode images used in the experiments.

Table 1

Description of simulated B-mode images used in the experiments; Number density of scatterers in the background: 5000 cm^{-3}.

Code name	Transducer frequency (MHz)	Lesion/background amplitude (dB)	Number density of scatterers in lesion (cm^{-3})
derh03f3	3	+3	5000
derh05f3	3	+5	5000
derl03f3	3	−3	5000
derl05f3	3	−5	5000
dlrh03f3	3	+3	500
dlrl03f3	3	−3	500
derh03f5	5	+3	5000
derl03f5	5	−3	5000

Figures 1a and 1b depict simulated B-mode images where the lesion scatter strength is 3 dB above that of the background (hyperechoic lesions). The lesion and the background in Fig. 1a have equal number densities of scatterers. The number density of scatterers in the lesion area in Fig. 1b is 10 times lower than that of the surrounding area. Both images were recorded with a 3 MHz transducer. The images have dimensions 241 × 241 and a resolution of 6 bits/pixel.

The gray level histograms of the pixels belonging to the lesion area and to the background for both images are plotted in Figs. 2a and 2b, respectively.

Patterns can be collected by running a window over the simulated ultrasound B-mode images. The patterns comprise pixel gray levels within the running window. Since the lesion area is well defined, we can easily have the ground truth for each pattern, if we agree that when the central pixel of the running
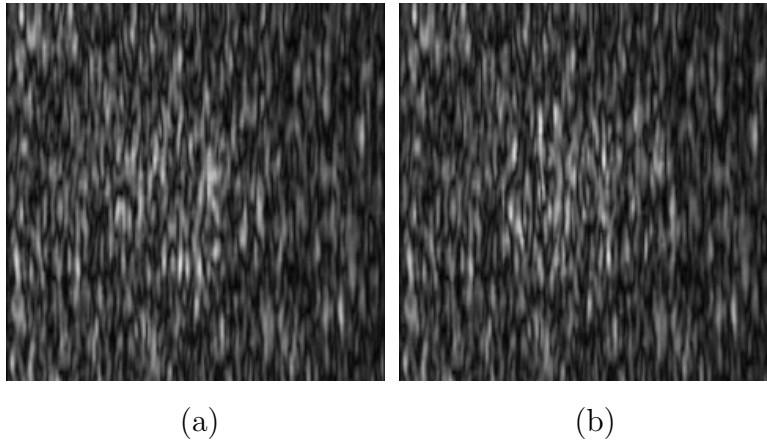
(a)   (b)

Fig. 1. Simulated ultrasound B-mode images: (a) original image derh03f3; (b) original image dlrh03f3. (The original gray levels are multiplied by 4 for display purposes.)
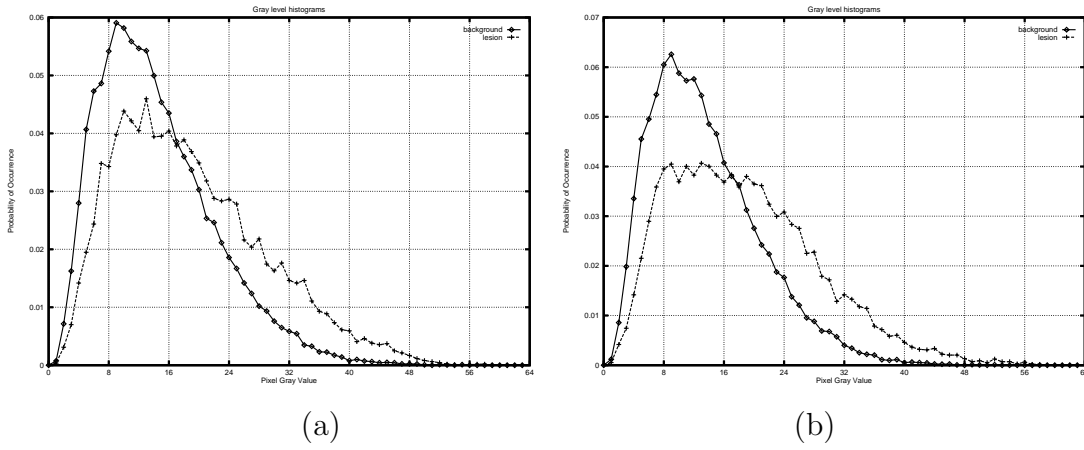

(a)   (b)

Fig. 2. Gray level histograms of the pixels belonging to the lesion and background areas for the (a) original image derh03f3; (b) original image dlrh03f3.

window falls inside the circular lesion then the training pattern is labeled as a positive one ($y_i = +1$); otherwise, it is labeled as a negative one ($y_i = -1$). The total number of patterns depends on the shape and the dimensions of the running window. The shape of the window is recommended to be related to the lateral and axial correlation size for the ultrasound B-mode image used. For example, for the images enlisted in Table 1 the speckle dimensions are 15 pixels in the lateral (vertical) direction and 3 pixels in the axial (horizontal) direction (Verhoeven et al., 1991). The experiments we conducted did not reveal any influence of the shape of the running window on the segmentation accuracy. Accordingly we decided to employ square windows. For 15×15 square windows, the total number of positive and negative patterns is 11289 and 40240, respectively. 1128 positive and another 1128 negative patterns were randomly selected to form the training set. The components of the training/test patterns are scaled by the maximum value they admit. The SVM^{light} implementation of SVMs was used in all experiments (Joachims, 1998).

11

To select the window dimensions and the most efficient kernel we used as figure of merit the area under the Receiver Operating Characteristic (ROC) measured on the test patterns. The ROC is defined here as the plot of the probability of detection versus the probability of false alarm. The area under the ROC corresponds to the fraction of the correct enforced answers when we detect the lesion in two stimuli and we know that the signal exists only in one of them (Thijssen, 1988). An SVM implementation usually returns a single operating point in the test phase. For example, SVM^{light} returns the pair of precision and recall. These probabilities are defined as follows. Precision is the proportion of image pixels detected as lesion pixels that are indeed lesion pixels, i.e.

$$P_{\text{precision}} = P\{y = 1 | f(\mathbf{x}) > 0\}. \tag{26}$$

Recall is the proportion of lesion pixels that are detected as lesion pixels, i.e.

$$P_{\text{recall}} = P\{f(\mathbf{x}) > 0 | y = 1\}. \tag{27}$$

To derive the ROC of SVM we compute the following pair:

$$P_{\text{precision}}(t) = P\{y = 1 | f(\mathbf{x}) > t\} \tag{28}$$
$$P_{\text{recall}}(t) = P\{f(\mathbf{x}) > t | y = 1\} \tag{29}$$

for various threshold selections t. It can easily be seen that $P_D(t) = P_{\text{recall}}(t)$ while $P_F(t)$ can be computed by

$$P_F(t) = \frac{P\{y = 1\}}{P\{y = -1\}} \left[\frac{1 - P_{\text{precision}}(t)}{P_{\text{precision}}(t)} \right] P_{\text{recall}}(t). \tag{30}$$

Figure 3 depicts the ROC curves of SVMs with several kernels and window dimensions trained on the simulated ultrasound B-mode image derh03f3. The training set was formed by 2256 patterns, as described previously. The test set was the total number of patterns collected by running a square window of the corresponding dimensions over the same image. The penalty term C in (19) was set to the default value by the SVM^{light}. The particular kernel functions used are tabulated in Table 2. The parameter values of the kernels enlisted in Table 2 resulted in the best accuracy on the test set experiments. The ROC curve for the L_2 mean (Kotropoulos and Pitas, 1992) is overlaid in all figures for comparison purposes. It is seen that the RBF kernel defined in (25) with parameter $\gamma = 2$ yields a better ROC curve than that of the other kernels. Moreover, the area under the ROC curve becomes greater as the window dimensions increase. We decided to fix the window dimensions to
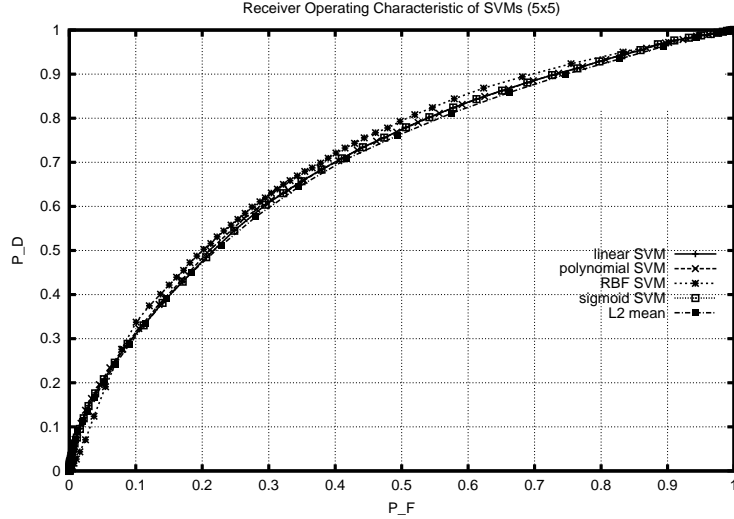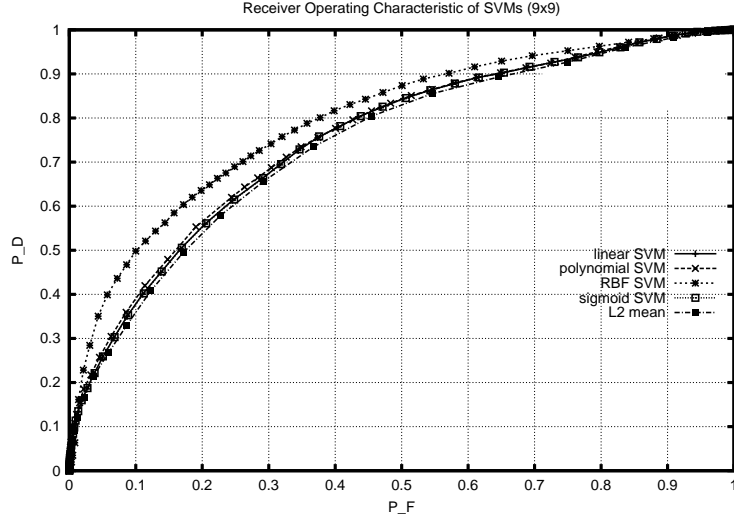
Fig. 3. ROC curves of SVMs for various kernels. (a) Running window of dimensions 5×5; (b) Running window of dimensions 9×9.

15×15 that roughly corresponds to 1×5 the speckle dimension in the lateral and axial directions, respectively.

Having fixed the window dimensions to 15×15 and chosen the RBF kernel function we proceed to the study of the selection of the parameter γ in (25) and the penalty factor C in (19). The most used choice for γ and C is that corresponding to the minimum of the test error, i.e., the error measured on the test data. Alternatively, the parameters can be chosen to minimize an upper bound of the expected error as is measured on the training set by applying a leave-one-out cross-validation procedure. Two such bounds are derived in (Vapnik, 1998; Chapelle and Vapnik, 2000). In the framework of leave-one-out

Table 2
Kernel functions employed in SVMs for the segmentation of the simulated ultrasound B-mode image derh03f3.

Kernel	$K(\mathbf{x}_i, \mathbf{x}_j)$
Linear	$\mathbf{x}_i^T \mathbf{x}_j$
Polynomial	$(\mathbf{x}_i^T \mathbf{x}_j + 2.0)^2$
RBF	$\exp(-2.0 \, \|\mathbf{x}_i - \mathbf{x}_j\|^2$
Sigmoid	$\tanh(0.001 \, \mathbf{x}_i^T \mathbf{x}_j + 0.1)$

estimates a numerically attractive solution for the automatic derivation of the best parameters γ and C within a predetermined range has been described in (Lee and Lin, 2001). The latter method has been employed in the RBF-SVM used to segment the simulated ultrasound B-mode image derh03f3. The search space was made by the choices $C \in \{1, 2, 4, 8, 16\}$ and $\gamma \in \{0.5, 1, 2, 4, 8\}$. The best model parameters were $\gamma = 1$ and $C = 8$. Table 3 summarizes the leave-one-out accuracy measured on the training set, the test accuracy, and the area under the ROC curve. For completeness, the margin, the number of the support vectors, and the estimated upper bound on the VC dimension are included in Table 3. The entries for the ad-hoc choice $\gamma = 2.0$ and $C = 0.5$ are also given for comparison purposes. In the following, we shall employ the optimal parameters for γ and C determined by the procedure described previously.

Table 3
Selection of model parameters in an RBF-SVM used to segment the simulated ultrasound B-mode image derh03f3 (Filter window: 15×15).

Parameters/ method	Leave-one-out accuracy (%)	Test accuracy (%)	Area under the ROC	Margin	Number of SVs	Upper bound on VC dimension
$\gamma = 1.0$, $C = 8.0$ (Lee and Lin, 2001)	94.105	80.12	0.867566	31.461	1650	1981
$\gamma = 2.0$, $C = 0.5$ (ad-hoc)	83.777	80.77	0.852047	26.985	2130	1407

Next we compared the performance of an RBF-SVM to that of the method proposed by Collaris and Hoeks (Collaris and Hoeks, 1995) and the thresholded L_2 mean filtered image (Kotropoulos and Pitas, 1992) on the simulated ultrasound B-mode image derh03f3. An additional figure of merit taken into consideration is the probability of detection P_D for a fixed probability of false alarm $P_F \in [5\%, 10\%]$. To enable comparisons we shall use as reference the

14

probability of false alarm attained when we threshold the L_2 mean filtered image by a suitable threshold. As has been noted, a comparison based on the probability of detection P_D for a fixed probability of false alarm P_F holds only for this operating point and is sometimes an inadequate figure of merit, especially when ROC curves cross (Wagner, 1986). On the contrary, the area under the ROC is a more reliable figure of merit for a complete characterization of the whole ROC curve. The choice of L_2 mean is not a trivial one, because the L_2 mean filter was proven to be the maximum likelihood estimator (ML) of a constant signal corrupted by a multiplicative Rayleigh speckle noise (Kotropoulos and Pitas, 1992). Moreover, the ML estimator for Gaussian signal-dependent speckle was shown that it resembled the L_2 mean filter and led to the L_2 variant of the self-organizing map (Kotropoulos et al., 1994). Table 4 summarizes the figures of merit measured on the test patterns. The figures quoted for the method developed in (Collaris and Hoeks, 1995) are the best ones. They have been obtained by setting both parameters q and K to 1.0. It is seen that the RBF-SVM outperforms its competitors. In the following the thresholded L_2 mean filtered image is employed as baseline technique in all comparisons.

Table 4
Figures of merit for lesion detection on the simulated ultrasound B-mode image derh03f3 by applying several methods.

Method	Area under the ROC	P_F (%)	P_D (%)
(Collaris and Hoeks, 1995)	0.680	6.61	22.78
L_2 mean	0.796	6.52	42.50
RBF-SVM	0.867	5.99	55.23

Table 5 summarizes the figures of merit for lesion detection on the simulated ultrasound B-mode images described in Table 1 for an RBF-SVM that was trained on 2256 training vectors of dimensions (225×1) using the optimal parameters for γ and C extracted from the same image. It is seen that the SVM with an RBF kernel outperforms the baseline technique in all cases. The CPU time needed to segment a simulated ultrasound B-mode image of dimensions 241×241 was approximately 1300 sec on a SUN Ultra-10 workstation with 256 MB RAM. The training procedure lasted approximately 34 sec in all cases.

Figure 4 depicts the ROC curves for two ultrasound B-mode images. In these images the lesion and the background have the same number density of scatterers and the lesion/background amplitude is +3 dB. The transducer frequency varies also from 3 MHz to 5 MHz. To assess the statistical significance of

Table 5
Figures of merit for lesion detection on simulated ultrasound B-mode images (Filter window: 15×15).

Code	RBF-SVM					L_2 mean		
name	γ	C	Area under ROC	P_F (%)	P_D (%)	Area under ROC	P_F (%)	P_D (%)
derh03f3	1	8	0.867	5.99	55.23	0.796	6.52	42.50
derh05f3	1	2	0.947	9.55	86.56	0.916	9.77	78.37
derl03f3	2	4	0.878	8.34	63.34	0.826	8.70	47.61
derl05f3	2	16	0.921	9.28	81.65	0.905	8.80	73.37
dlrh03f3	1	2	0.919	7.09	69.17	0.855	6.86	52.37
dlrl03f3	2	4	0.838	7.63	52.26	0.792	8.79	39.81
derh03f5	2	1	0.903	5.74	78.03	0.855	8.40	63.02
derl03f5	2	1	0.976	9.64	94.56	0.952	10.88	88.74

the rates observed, we model the ensemble {test patterns, segmentation algorithm} as a source of binary events, 1 for a false alarm or 0 for a correct rejection, with probability P_F of drawing a 1 and $(1 - P_F)$ of drawing a 0. Similarly another source of binary events, 1 for correct detection or 0 for a false detection, with probability P_D of drawing a 1 and $(1 - P_D)$ of drawing a 0 is also assumed. These events can be described by Bernoulli trials. Let us denote by $\hat{P}_F$ and $\hat{P}_D$ the estimates of P_F and P_D, respectively. The exact ϵ confidence interval of P_F is the segment between the two roots of the quadratic equation (Papoulis, 1991):

$$(P_F - \hat{P}_F)^2 = \frac{z_{(1+\epsilon)/2}^2}{40240} \, P_F \, (1 - P_F) \tag{31}$$

where z_u is the u-percentile of the standard Gaussian distribution having zero mean and unit variance. In a similar manner, the exact ϵ confidence interval of P_D is the segment between the two roots of the quadratic equation:
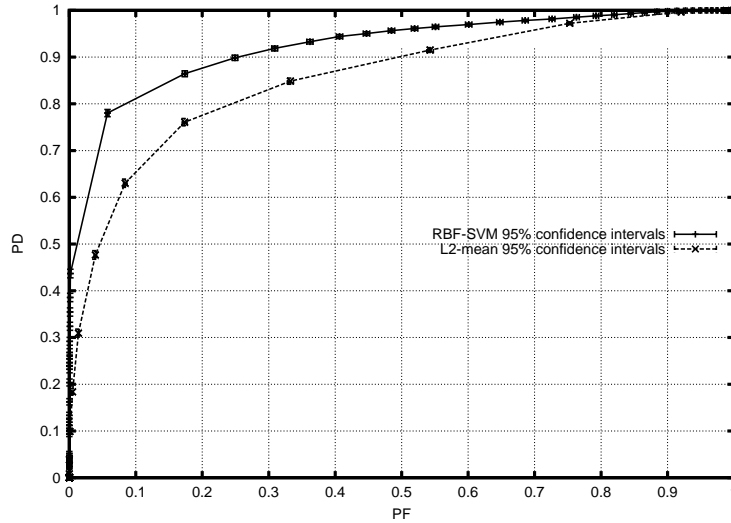
$$(P_D - \hat{P}_D)^2 = \frac{z_{(1+\epsilon)/2}^2}{11289} \, P_D \, (1 - P_D). \tag{32}$$

The $\epsilon = 95\%$ confidence interval of P_F and P_D is indicated with a horizontal and a vertical error bar, respectively, in the ROCs of Fig. 4. As can be seen the improvements reported are statistically significant.

Moreover, the trained RBF-SVMs on the aforementioned set of tissue mimick-
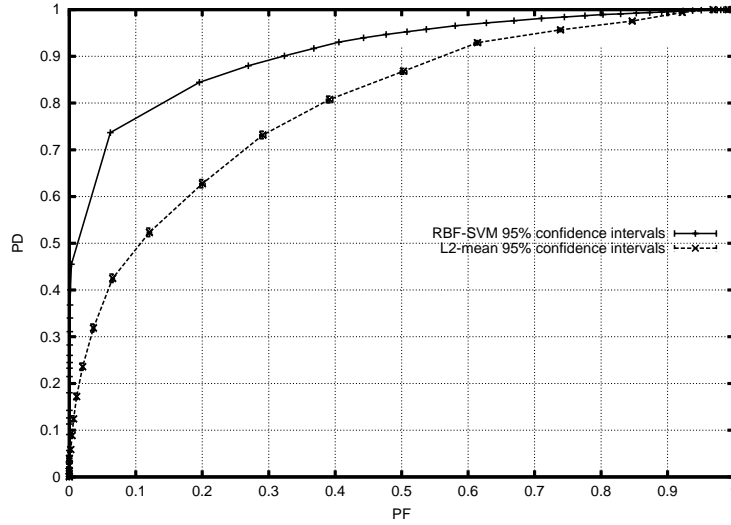
(a)

(b)

Fig. 4. ROC curves of RBF-SVM and L_2 mean for two simulated ultrasound B-mode images (a) derh03f3; (b) derh03f5.

ing ultrasound B-mode images were successfully applied on a different set of B-mode images from a contrast/detail phantom recorded at the Biophysics Laboratory, Department of Ophthalmology, University Hospital Nijmegen, The Netherlands. Each image contains a region of interest that is displayed in sector format. In each sector two lesions are present, although they are not always visible. The images were obtained by scanning a contrast-detailed phantom, whose description can be found in (Smith et al., 1983), in such a way that the cones were intersected perpendicularly by the ultrasound beam. Two images are considered that contain two lesion cross-sections of a large diameter and a contrast of 3.6 dB and 7.1 dB relative to the background, respectively. These images are shown in Fig. 5a and 5d. To segment the le-

sions, we tested RBF-SVMs trained on ultrasound B-mode images derh03f3 and derh05f3 with patterns of dimensions (9×1) or (25×1). The window size was chosen so that the fine details of the images are preserved. The best results were obtained by using the trained RBF-SVM with window size 3×3 on derh05f3. The segmented lesions are shown with white gray level in Fig. 5b and 5e, respectively. Based on visual judgements, it can be seen that the RBF-SVM segments the lesion satisfactorily in both test images. For comparison purposes, the segmentations produced by thresholding the L_2 mean filtered images are depicted in Fig. 5c and 5f, respectively. The thresholds have been chosen so that the same number of positive and negative test examples to that produced by the RBF-SVM results. It is seen that the two segmentation results correlate. However, no threshold selection is needed for the RBF-SVM.

(a)　　　　　　　　(b)　　　　　　　　(c)
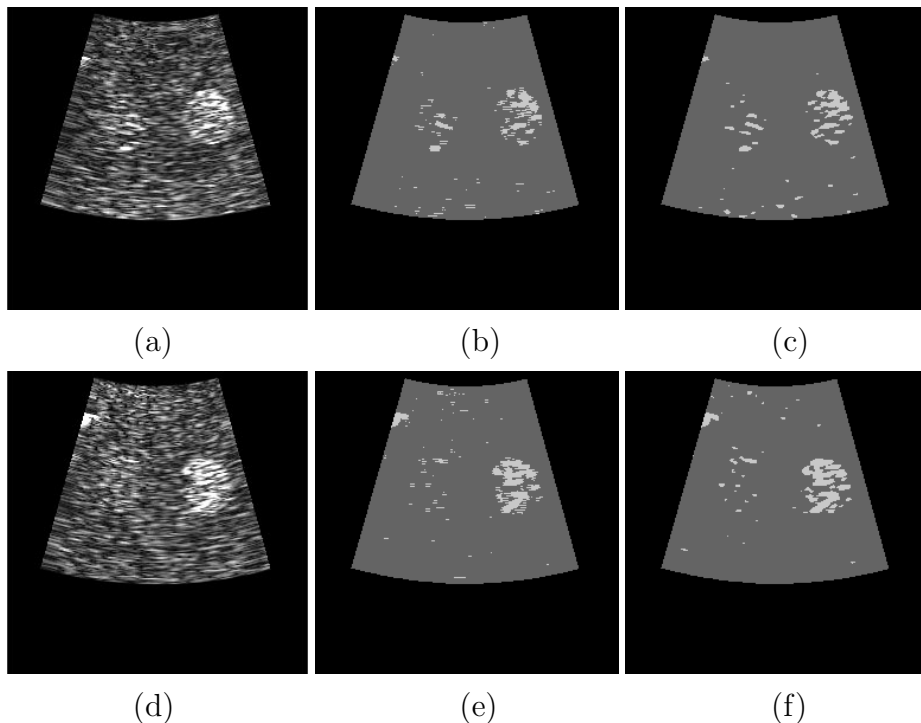
(d)　　　　　　　　(e)　　　　　　　　(f)

Fig. 5. (a) Original image depicting two lesions obtained from cone cross-sections of large diameter and mid-light reflection amplitude. (The original amplitudes are multiplied by 3 for display purposes.) (b) Segmentation into two classes using a trained RBF-SVM on the tissue mimicking phantom depicting a circular lesion that has reflection strength 5 dB higher than the background and equal number density of scatterers with the background (derh05f3). (c) Segmentation derived by thresholding the L_2 mean filtered original image (a). (d) Original image depicting two lesions obtained from cone cross-sections of large diameter and mid-light reflection amplitude. (The original amplitudes are multiplied by 3 for display purposes.) (e) Segmentation result using the same RBF-SVM as in (b). (f) Segmentation derived by thresholding the L_2 mean filtered original image (d).

Finally, we applied the trained RBF-SVM on ultrasound B-mode image derh05f3 with patterns of dimensions (9×1) or (25×1) to a clinical ultrasonic image of liver. The image was recorded with a 3 MHz probe at the Institute of Cancer Research of the Royal Marsden Hospital, Sutton, U.K. The original image, its segmentation into two classes using the RBF-SVM, and the segmentation produced by thresholding the L_2 mean filtered original image are shown in Fig. 6a, 6b, and 6c, respectively. The threshold was chosen as described previously. Again, based on visual judgements the segmentation is found satisfactory.
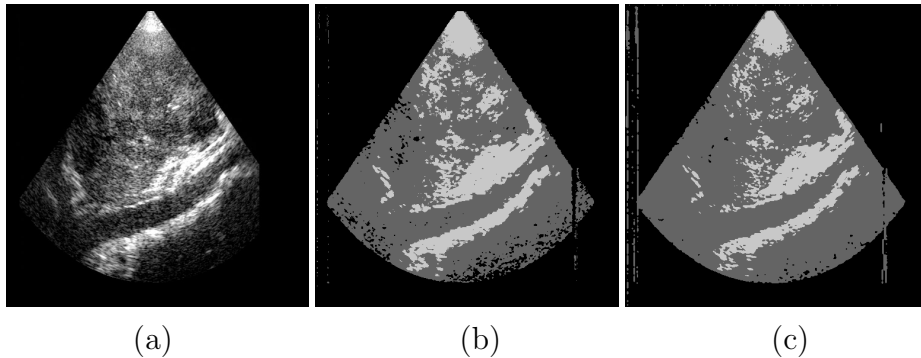

(a)   (b)   (c)

Fig. 6. (a) Original image of liver recorded using a 3 MHz probe. (The original gray level values are multiplied by 2 for display purposes.) (b) Segmentation into two classes using the same RBF-SVM as in Fig. 5. (c) Segmentation derived by thresholding the L_2 mean filtered original image (a).

4 Conclusions

In this paper we have proposed the use of SVMs to segment lesions in ultrasound images. We have demonstrated by experiments that RBF-SVMs outperform the thresholding of L_2 mean filtered images under several lesion and recording conditions in ultrasound B-mode images. Moreover, we have found that the trained RBF-SVMs segment satisfactorily other (unseen) ultrasound B-mode images and clinical ultrasonic images.

References

Ashton, E., Parker, K., 1995. Multiple resolution bayesian segmentation of ultrasound images. Ultrasonic Imaging 17, 291–304.

Bonneville, M., Meunier, J., Bengio, Y., Soucy, J.-P., 1998. Support vector machines for improving the classification of brain pet images. In: Hanson, K. M. (Ed.), Proc. SPIE Medical Imaging. Vol. 3338. pp. 264–273.

Burges, C., 1998. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2 (2), 1–43.

Caves, R., Quegan, S., White, R., November 1998. Quantitative comparison of the performance of SAR segmentation algorithms. IEEE Transactions on Image Processing 7 (11), 1534–1546.

Chapelle, O., Vapnik, V., 2000. Model selection for support vector machines. In: Solla, S., Leen, T., Müller, K.-R. (Eds.), Advances in Neural Information Processing Systems. Vol. 12. The MIT Press, Cambridge, MA.

Collaris, R., Hoeks, A., 1995. Improved lesion detection by level-dependent spatial summation. Ultrasonic Imaging 17, 197–226.

Cristianini, N., Shawe-Taylor, J., 2000. An introduction to support vector machines. Cambridge University Press, Cambridge, U.K.

Dutt, V., Greenleaf, J., December 1996. Adaptive speckle reduction filter for log-compressed B-scan images. IEEE Transactions on Medical Imaging 15 (6), 802–813.

Evgeniou, T., Pontil, M., Papageorgiou, C., Poggio, T., 2000a. Image representations for object detection using kernel classifiers. In: Proc. Asian Conf. Computer Vision. pp. 687–692.

Evgeniou, T., Pontil, M., Poggio, T., 2000b. Regularization networks and support vector machines. In: Smola, A., Bartlett, P., Schölkopf, B., Schuurmans, D. (Eds.), Advances in Large Margin Classifiers. The MIT Press, Cambridge, MA, pp. 171–204.

Fletcher, R., 1987. Practical Methods of Optimization, 2nd ed. John Wiley, New York.

Joachims, T., 1998. Making large-scale SVM learning practical. Technical Report LS8-Report 24, Universität Dortmund.

Karaman, M., Kutay, M. A., Bozdagi, G., June 1995. An adaptive speckle suppression filter for medical ultrasonic imaging. IEEE Transactions on Medical Imaging 14 (2), 283–292.

Klingler, J., Vaughan, C., Fraker, T., Andrews, L., November 1988. Segmentation of echocardiographic images using mathematical morphology. IEEE Trans. on Medical Imaging 35 (11), 925–934.

Kotropoulos, C., Magnesalis, X., Pitas, I., Strintzis, M., January 1994. Nonlinear ultrasonic image processing based on signal-adaptive filters and self-organizing neural networks. IEEE Transactions on Image Processing 3 (1), 65–77.

Kotropoulos, C., Pitas, I., 1992. Optimum nonlinear signal detection and estimation in the presence of ultrasonic speckle. Ultrasonic Imaging 14 (3),

249–275.

Lee, J.-H., Lin, C.-J., 2001. Automatic model selection for support vector machines. Machine Learning To appear.

Mangasarian, O., Street, W. N., Wolberg, W., July-August 1995. Breast cancer diagnosis and prognosis via linear programming. Operations Research 43 (4), 570–577.

Mohan, A., Papageorgiou, C., Poggio, T., April 2001. Example-based object detection in images by components. IEEE Trans. Pattern Analysis and Machine Intelligence 23 (4), 349–361.

Müller, K.-R., Mika, S., Rätsch, G., Tsuda, K., Schölkopf, B., March 2001. An introduction to kernel-based learning algorithms. IEEE Transactions on Neural Networks 12 (2), 181–201.

Muzzolini, R., Yang, Y., Pierson, R., March 1993. Multiresolution texture segmentation with application to diagnostic ultrasound images. IEEE Transactions on Medical Imaging 12 (1), 108–123.

Osuna, E., Freund, R., Girosi, F., Feb. 28, 1997. Support vector machines: Training and applications. Technical Memo AIM-1602, Massachusetts Institute of Technology, Artificial Intelligence Laboratory.

Papageorgiou, C., Poggio, T., September 1999. A pattern classification approach to dynamic object detection. In: Proc. Int. Conf. Computer Vision. Vol. 2. pp. 1223–1228.

Papoulis, A., 1991. Probability, Random Variables and Stochastic Processes, 3rd Edition. McGraw-Hill, N.Y.

Rekeczky, C., Tahy, A., Vegh, Z., Roska, T., January-February 1999. Spatio-temporal nonlinear filtering and endocardial boundary detection in echocardiography. Int. Journal of Circuit Theory and Applications 27, 171–207.

Schoelkopf, B., Sung, K., Burges, C., Girosi, F., Niyogi, P., Poggio, T., Vapnik, V., November 1997. Comparing support vector machines with gaussian kernels to radial basis function classifiers. IEEE Trans. on Signal Processing 45 (11), 2758–2765.

Shawe-Taylor, J., Bartlett, P., Williamson, R., Anthony, M., September 1998. Structural risk minimization over data-dependent hierarchies. IEEE Trans. Information Theory 44 (5), 1926–1940.

Smith, S., Wagner, R., Sandrik, J., Lopez, H., May 1983. Low contrast detectability and contrast/detail analysis in medical ultrasound. IEEE Transactions on Sonics and Ultrasonics 30 (3), 164–172.

Thijssen, J., 1988. Focal lesions in medical images: A detection problem. In: Viergever, M., Todd-Prakopek, A. (Eds.), Proceedings NATO-ASI on Mathematics and Computer Science in Medical Imaging. Springer, Berlin, pp. 415–440.

Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer Verlag, New York.

Vapnik, V., 1998. Statistical Learning Theory. J. Wiley, New York.

Verhoeven, J., Thijssen, J., Theeuves, A., 1991. Lesion detection by echographic image processing: signal-to-noise ratio imaging. Ultrasonic Imaging

13, 239–251.

Wagner, R., 1986. Fundamentals and applications of signal detection theory in medical imaging. In: Proc. SPIE. Vol. 626. pp. 761–765.