



Scale selection for supervised image segmentation[☆]

Yan Li ^{a,*}, David M.J. Tax ^a, Marco Loog ^{a,b}

^a Pattern Recognition Laboratory, Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

^b The Image Group, Department of Computer Science, University of Copenhagen, Denmark

ARTICLE INFO

Article history:

Received 23 June 2011

Received in revised form 20 August 2012

Accepted 22 August 2012

Keywords:

Scale selection

Image segmentation

Supervised learning

Scale space theory

Max rule

ABSTRACT

Finding the right scales for feature extraction is crucial for supervised image segmentation based on pixel classification. There are many scale selection methods in the literature; among them the one proposed by Lindeberg is widely used for image structures such as blobs, edges and ridges. Those schemes are usually unsupervised, as they do not take into account the actual segmentation problem at hand. In this paper, we consider the problem of selecting scales, which aims at an optimal discrimination between user-defined classes in the segmentation. We show the deficiency of the classical unsupervised scale selection paradigms and present a supervised alternative. In particular, the so-called max rule is proposed, which selects a scale for each pixel to have the largest confidence in the classification across the scales. In interpreting the classifier as a complex image filter, we can relate our approach back to Lindeberg's original proposal. In the experiments, the max rule is applied to artificial and real-world image segmentation tasks, which is shown to choose the right scales for different problems and lead to better segmentation results.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

We consider the problem of scale selection for supervised image segmentation, which not only divides an image into disjoint areas, but also assigns each region or pixel into one of the pre-defined classes [1–7]. Usually the classes are defined by giving a set of labeled images, based on which a classifier is trained to capture the characteristics of pixels in each class. The classifier is then applied to new test images and assigns a class label to each pixel. In contrast, segmenting images without class assignment [8,9] may be referred to as unsupervised segmentation.

To train a classifier for supervised segmentation, one needs to know the scales to extract features. However, it is usually not known a prior what the right scales are for a problem at hand. This is where scale selection comes in, which typically examines an image by representing it at multiple scales and chooses scales according to a particular criterion.

The most well-known multi-scale image representations are linear and nonlinear scale spaces [10–13], Laplacian and Gaussian pyramids [14], and wavelet theory [15]. Those representations are closely related. A Gaussian pyramid without downsampling at higher levels is a linear scale space, which represents different scales by applying different amount of blur to the original image. It is also shown that results in wavelet theory can be transferred into scale space theory with variable substitution [16]. We use scale space as it provides a convenient way to segment an image at multiple scales and to compare the classification of

a pixel across the scales. Linear scale space is widely applied to detect and locate simple image structures like edges, blobs, corners, and ridges in a scale-dependent way [17–19]. Similar approaches are applied as parts of more complicated and higher-level tasks such as image description techniques, matching, and tracking [20–23].

The most widely-used scale selection scheme is probably Lindeberg's approach [18,19], which serves as the basis of many of the aforementioned techniques. Its basic idea is to normalize a differential operator for an image structure and apply the operator at all scales. Subsequently, the responses of the normalized operator at all scales are compared and the scale at which the absolute value of the response achieves the maximum is selected as the local scale of the image structure. In a sense, the optimal scale is the one at which the structure is most pronounced, and most discernible from its surroundings.

Other scale selection schemes include [24] which selects a scale with the maximum entropy at each pixel, [25] for anisotropic scale space, [26] which selects the scale having the maximum likelihood of an image model, [27] which assigns a larger scale to a pixel if its intensity changes less when total variation regularization is applied to the image, [28] which finds the so-called most contrasted shape for each pixel and computes a scale from that shape, [29] which selects a global scale for an image by defining generalized entropies in scale spaces, and [30–32] which are specifically designed for textures.

The statistical properties of the same object at different scales are studied in [33]. It is shown that generative models such as wavelet sparse coding are appropriate for images taken at a near distance (objects at small scales), while descriptive models such as Markov random fields are appropriate for images taken at a far distance (objects at large scales). It is thus very important to estimate the right scale (or entropy rates in

[☆] This paper has been recommended for acceptance by Thomas Brox.

* Corresponding author. Tel.: +31 15 2788433; fax: +31 15 2781843.

E-mail addresses: yan.li@tudelft.nl (Y. Li), d.m.j.tax@tudelft.nl (D.M.J. Tax), m.loog@tudelft.nl (M. Loog).

practice) and choose the appropriate model when computing a symbolic representation of an image, such as the primal sketch model in [34].

Lindeberg's and related schemes deal with unsupervised scale selection in the sense that they propose generally applicable selection mechanisms. A supervised method, on the other hand, takes into account the specific segmentation problem one actually wants to solve and selects scales relevant to this task. For example, for the same image containing a person, small scales are needed if the problem is to segment the eyes, while large scales are needed if the problem is to segment the person.

Another added benefit of supervised methods is that we can construct, or rather, learn relatively complex filters [35,36] for various problems, instead of pre-designed differential operators such as the Laplacian. Scales are then selected based on the complex filters.

The basic idea of our method is to let the classifier itself select appropriate scales for the supervised segmentation problem. Specifically, the classifier outputs a confidence measure about its result, which is used to select final segmentation results and supervised scales. Our experiments show that compared with unsupervised scales, the supervised ones are more suitable for the target segmentation problem and thus lead to much better results.

This paper is a revised and extended version of [37], which studied scale selection from a classifier combining perspective. The current paper takes the perspective of image segmentation. The limitations of unsupervised scale selection schemes and the need for supervised scale selection are discussed, and the segmentation method based on Lindeberg's scheme is compared in our experiments. The authors' work in [38] also deals with supervised scales, but is different as it is designed for scale-invariant segmentation. When an image is zoomed in or out, the label image outputted from the classification should be the same, but zoomed in or out accordingly. A single classifier is trained for the segmentation, in comparison with one classifier per scale in the current paper.

The paper is organized as follows. Sections 2 and 3 introduce scale space and selection, and supervised image segmentation by pixel classification, respectively. Section 4 discusses the deficiency of unsupervised scale selection for supervised image segmentation, which motivates our supervised scale selection in Section 5. Our method is tested in Section 6, and its segmentation results are compared with those from unsupervised scale selection and from concatenated multi-scale features. Several related issues are discussed in Section 7 and finally, Section 8 concludes the paper.

2. Scale space, N-jet, and Lindeberg's scheme

Motivated both by the viewpoint of physical measurements [16] and by the theory of biological visual perception [11,39], scale space theory provides a formal way to approach the problem of multi-scale image analysis. One of the most-often used approaches is linear, or Gaussian, scale space, which relies on a physically plausible axiomatization [16]. This scale space is also the one that we consider.

2.1. Linear scale space

Given an image $\ell : \mathbb{R}^2 \rightarrow \mathbb{R}$, its multi-scale representation is constructed as a one-parameter family of blurred images. The blurring corresponds to a convolution with a Gaussian kernel, whose variance σ^2 is proportional to the scale that the original image is represented. That is, a scale space representation $L : \mathbb{R}^2 \times \mathbb{R}^+ \rightarrow \mathbb{R}$ of ℓ is given by

$$L(x, y; \sigma) = (\ell * g_\sigma)(x, y) \quad (1)$$

in which the Gaussian kernel g_σ is defined as

$$g_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}.$$

The scale space framework can be extended to a theory for Gaussian derivative operators in which the general k th order derivatives

$$L_{x^{k_1}y^{k_2}}(x, y; \sigma) = \frac{\partial^k}{\partial x^{k_1}\partial y^{k_2}} g_\sigma, \quad (2)$$

where $k_1 + k_2 = k$, are used to convolve the input image $\ell(x, y)$ [16,40,39]. It was shown in [40] that the set of convolution filters with order k from 0 to ∞ constructs a complete, hierarchically ordered family of scale-space kernels.

A basic collection of features is the so-called N -jet, which are all the Gaussian image derivatives up to order N at a particular scale. For example, the 2-jet of image ℓ at location (x, y) and scale σ is the concatenation of derivatives

$$[L, L_x, L_y, L_{x^2}, L_{y^2}, L_{xy}](x, y; \sigma). \quad (3)$$

It is these basic features that are often employed in supervised image analysis techniques (e.g., [41–43]). The collection of basic linear filters, in turn, is used as a basis for expressing a large class of more general, potentially nonlinear, image processing operators [40,19,18,13], e.g., differential invariants [44–46].

2.2. Lindeberg's scale selection scheme

As an image structure typically resides at a particular scale, it is crucial to select locally appropriate scales for a specific structure in an image. Lindeberg [18,19] proposes a general automatic scale selection framework for image structures which can be detected by differential operators. As an example, this framework is described with the Laplacian as the differential operator for blob detection. The Laplacian of an image $L(x, y; \sigma)$ is

$$\Delta L = L_{xx} + L_{yy}, \quad (4)$$

where the variables $(x, y; \sigma)$ are omitted for brevity.

To compensate for the amplitude decrease of spatial derivatives for larger scales, Lindeberg proposes to normalize the derivatives to make them comparable across the scales. The introduced normalization corresponds to multiplying a k th order derivative with σ^k (a more general form used in [18] is $(\sigma^k)^\gamma$). Since all the differential entities in the Laplacian are of order 2, its corresponding normalization is to multiply ΔL with σ^2 :

$$\Delta L_{\text{norm}} = \sigma^2 (L_{xx} + L_{yy}). \quad (5)$$

The normalized differential operator is applied to an image at all scales, and the scale at which the operator attains the largest response is selected as the local scale for the corresponding image structure. A simple interpretation of this idea is that the scale is selected at which the structure's presence is most pronounced or best matches the hypothesis. Consider for example the normalized Laplacian, which is applied to an image at all scales

$$\Delta L_{\text{norm}}(x, y; \sigma), \forall \text{ for all } \sigma \in \mathbb{R}^+. \quad (6)$$

At each location (x, y) , the scale is selected at which the response of the normalized Laplacian attains the maximum across scales

$$S_{\text{blob}}(x, y) = \underset{\sigma \in \mathbb{R}^+}{\operatorname{argmax}} \Delta L_{\text{norm}}(x, y; \sigma). \quad (7)$$

Lindeberg's scheme has been widely used in computer vision and image processing. Together with the Laplacian or the determinant-of-Hessian $L_{xx}L_{yy} - L_{xy}^2$ as a blob detector, it is employed in image keypoint detection and description such as Harris–Laplace detector

[21], SIFT [20] and SURF [47]. These structure detectors and descriptors are often used in image matching, tracking, recognition, etc. [47,23,48,20,49,21]. Other applications of Lindeberg's scheme include [22], which normalizes scale variations of objects for better recognition, [50], which achieves scale invariant detection and recognition, [51], which selects scales of blobs corresponding to words, and [24,52], which select scales for saliency measures.

3. Supervised image segmentation by pixel classification

Pixel classification is one of the common ways to perform supervised image segmentation. It can adapt well to different segmentation tasks with different training data and is able to provide state-of-the-art performance in various settings [53–55,7,42,41,43,5,56]. Its basic idea is to extract a feature vector $F(x,y)$ for each pixel, and to train a classifier h for obtaining labels of pixels.

The term feature vector is used as in pattern recognition and machine learning and refers to the measurements describing a particular object. In our case, this typically means image measurements or descriptors that provide low-level information related to every position in the image. Examples are gradients, N -jets, responses of Gabor filters, and Hessians related to every pixel. In principle, however, higher-level descriptors like locally binary patterns, histograms of gradients, and SIFT would also be possible. The N -jets are used as features in our experiments.

The information about different classes of pixels is contained in the so-called training set, which contains n images $\ell_j, j = 1, \dots, n$ and their corresponding outputs c_j (called the label images). The output c_j takes value from a set of class labels, e.g. $\{1, \dots, K\}$ for a K -class problem. The label $c_j(x,y)$ denotes to which class the pixel $\ell_j(x,y)$ belongs.

The task of classifier training is to learn the characteristics of pixels of different classes from the training set. The trained classifier is a function $h : \mathbb{R}^d \rightarrow \{1, \dots, K\}$, which maps feature vectors $F(x,y)$ (of dimensionality d) to their class labels $c(x,y)$. Numerous classifiers have been proposed in the field of pattern recognition and machine learning [57,58], such as neural networks, the k -nearest neighbor method, quadratic discriminant analysis (QDA), and support vector machines.

3.1. Quadratic discriminant analysis (QDA) as the classifier

In our experiments QDA is used, which is a relatively simple classifier. It is based on the assumption that the feature vectors F from each class $\omega_k, k = 1, \dots, K$ follow a Gaussian distribution $g_k(F|\mu_k, \Sigma_k)$, where μ_k is the mean and Σ_k is the covariance matrix of the Gaussian. With π_k as the prior of class ω_k , the posterior probability of a feature vector F can be computed as

$$P(\omega_k|F) = \frac{g_k(F|\omega_k)\pi_k}{p(F)} \propto g_k(F|\omega_k)\pi_k. \quad (8)$$

According to the Bayes decision theory, a feature vector F is classified to the class having the largest posterior probability $P(\omega_k|F)$, or equivalently $g_k(F|\omega_k)\pi_k$. Substituting the Gaussian distribution $g_k(F|\omega_k)$, the decision may be based on the so-called quadratic discriminant functions

$$\delta_k(F) = -\frac{1}{2}(F - \mu_k)^T \Sigma_k^{-1} (F - \mu_k) - \frac{1}{2} \log |\Sigma_k| + \log \pi_k. \quad (9)$$

A feature F is classified to class ω_k if $\delta_k(F)$ is the maximum among all classes. The unknown parameters, μ_k, Σ_k and π_k , are estimated from the training set, and denoted as $\hat{\mu}_k, \hat{\Sigma}_k$ and $\hat{\pi}_k$. Substituting the

estimated parameters in Eq. (9), we can represent the training classifier as

$$h(F) = \underset{k \in \{1 \dots K\}}{\operatorname{argmax}} \delta_k(F; \hat{\mu}_k, \hat{\Sigma}_k, \hat{\pi}_k). \quad (10)$$

In the testing phase of classification, features F_t are extracted for all the pixels of a new test image ℓ_t and classified with the trained classifier h . The output $\hat{c}(x,y) = h(F(x,y))$ from the classifier is the estimated class label for $\ell_t(x,y)$ at the location (x,y) .

A pixel is misclassified if its estimated label $\hat{c}(x,y)$ is different from its true label $c(x,y)$. The error rate for the test image $\ell_t(x,y)$ is the fraction of misclassified pixels with respect to the total number of pixels.

3.2. Classification with Lindeberg's scheme

The feature vector of a pixel is usually extracted from a neighborhood of the pixel, and a crucial parameter is the size, or the scale, of the neighborhood. The widely-used Lindeberg's scheme may also be employed here to select scales. The idea is to select a scale $S(x,y)$ at each location (x,y) with a differential operator, and then extract a feature vector at that specific scale

$$F(x,y; \sigma = S(x,y)). \quad (11)$$

For example, when 2-jet is used as feature, it is extracted at the selected scale $S(x,y)$

$$[L, L_x, L_y, L_{x^2}, L_{y^2}, L_{xy}] (x, y; S(x, y)). \quad (12)$$

Compared with the general form Eq. (3), the scale σ to extract features has now been fixed to $S(x,y)$.

A normalized blob detector is employed as the differential operator in this paper. It is successfully used in keypoint detection and description for general images [20,21,47], showing that its selected scales reflect the scale well of the local image structure. It is also used in [22] to select a local scale at each location, with which the receptive field to compute features is normalized. In our experiments, the Laplacian is used as the blob detector.

4. The need for supervised scale selection

The unsupervised scales selected with Lindeberg's scheme are the characteristic sizes for the image structures, but they are not necessarily the best scales to be employed in a specific classification problem. In the following, we describe this difference in detail and provide two illustrations.

4.1. The problem with unsupervised scales

Lindeberg's scheme concerns the right scale to extract a specific image structure, such as a blob, a corner, an edge, or a ridge. Intuitively, the structure is most prominent at the selected scale and becomes less significant at other scales. Thus, the scale is the *characteristic length* of an image structure [18,21] and depends only on the size of the structure.

Different tasks, however, may call for different scales to be used. In the particular setting of supervised segmentation, one would like to select the scale at which classification can be done most reliably. Therefore, a natural definition of scale would be the one at which the discrimination between classes is maximized, potentially resulting in a lower segmentation error than when an unsupervised scheme is employed.

4.2. Illustrative examples

The following two brief experiments demonstrate that unsupervised scales cannot be optimal for supervised segmentation problems in

general. Unsupervised scales were selected with a normalized Laplacian. For these experiments, images were of size 256×256 and a total of 20 scales were sampled exponentially from [0.5, 200]. The QDA was used as the classifier and 2-jet and 4-jet as features for the two problems, respectively. The training set consisted of 5 images.

For comparison, the results of our supervised approach are also provided. It trains a classifier at each scale and selects the scale having the maximum posterior across the scale. It will be explained in full detail in the next section. The scale range was [0.5, 32] since the classifiers trained at scales larger than 32 did not add much information. The other experimental setting was the same as those for the unsupervised scale selection described in previous paragraph.

4.2.1. Disk segmentation

The first example that we consider is segmenting a bright disk in the center from a dark background. Though the Laplacian is an appropriate differential operator for blob detection, classification with the unsupervised scales still does not work well in some situations.

The first row of Fig. 1 shows the results on a test image without adding any noise. This is a very simple problem since the disk can easily be segmented at many scales, though too much blurring at a large scale would make the boundary locations less accurate. The classification with Lindeberg's scheme achieves a good result, and the unsupervised scales indicate the sizes of the blobs detected by the Laplacian. Note that the four corners of the image have large scales,

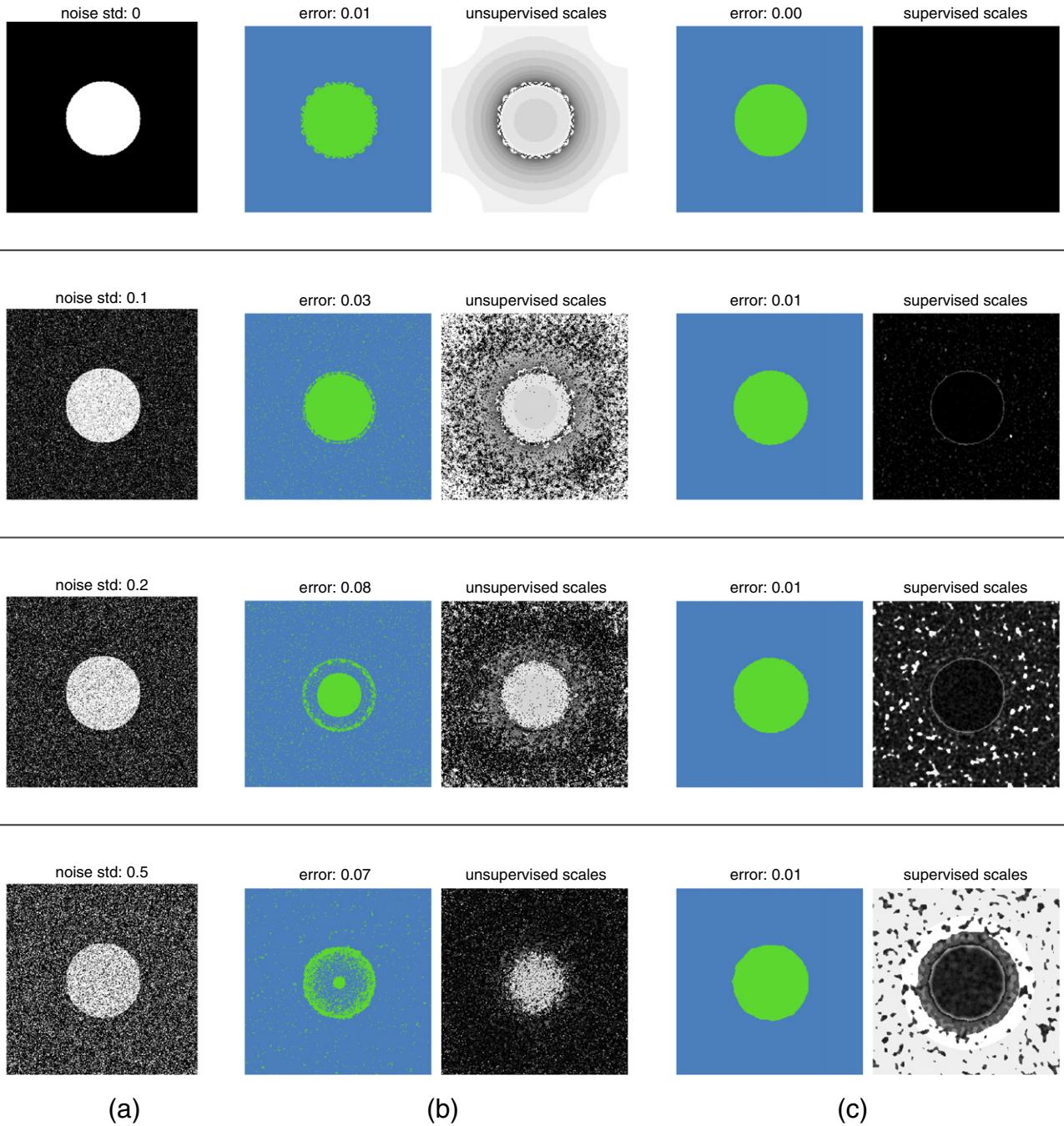


Fig. 1. Segmentation of a disk from its background. (a) The test images, contaminated with different amount of noise. They have identical label images (not shown here), which contain the disk in the center as one class and the background as the other class. (b) Results of classification with Lindeberg's scheme. (c) Results with our method, which is detailed in Section 5.1. In b and c, the left images are the segmentation results by pixel classification, and the right images show the log of the selected scales for better visuality (dark denotes small scales).

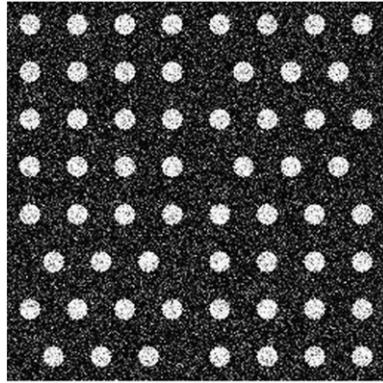


Fig. 2. An example image of the data set, which is a set of bright disks in the dark background. The disks are arranged in two lattices, rectangular and triangular. Gaussian noise with zero mean and standard deviation 0.1 is added.

since they are in a large homogeneous area, which is treated as a large blob by the Laplacian.

The second to the fourth rows of Fig. 1 show the result on the test image with different amount of Gaussian noise. The means of the noise are all zero and the standard deviations are indicated at the top of each subimage in Fig. 1(a). For noisy images, the classification with Lindeberg's scheme becomes less satisfactory: the disk in the center is not segmented well from the background and the results are quite noisy. As a certain amount of blurring is needed to segment the disk accurately in a noisy image, scale selection has a bigger influence on the classification than the previous problem dealing with noiseless images. The degraded results show that the unsupervised scales do not reflect well the scales needed for classification.

Fig. 1(c) shows the classification results with the supervised scales by our proposed method. For the disk in the center, the supervised scales are much smaller than the unsupervised scales, which means that the disk is best segmented at small scales. For the background, increasing the noise level leads to larger supervised scales, since more

blurring is needed to discriminate the background pixels correctly. In the first row of Fig. 1(c), the image of supervised scales is black, since no blurring is needed for this noiseless image and thus small scales are selected for all pixels. With the supervised scales, good classifications are obtained for images with all levels of noise.

4.2.2. Two textures

Our second example illustrates the necessity of taking into account the classification problem when selecting scales, which cannot be achieved with unsupervised scale selection methods.

Each image in this datasets is a collection of bright disks against a dark background, and the disks are arranged in two different lattices. Gaussian noise is added to each image separately. An example image is shown in Fig. 2.

We consider two different problems, which have the same input image, Fig. 2, but different label images, shown in Fig. 3(a). The first problem concerns segmentation between the disks and the background, and the second one is between the two lattices of disk locations.

Fig. 3(b) shows the classification results with Lindeberg's scheme. Identical unsupervised scales are selected for both problems, since the segmentation problems are not taken into consideration for scale selection. For the first problem of segmenting the disks from the background, the selected scales reflect the sizes of the disks and reasonable results are obtained. For the second problem of discriminating the two lattices, the classifier cannot capture the problem well with the unsupervised scales.

Fig. 3(c) shows the results with our method. Quite different scales are selected for the two problems. For the first problem, small scales are selected for disks and large ones for background, as more blurring is needed to segment the background reliably. For the second problem, the supervised scales are block-wise, showing that the scales required for pixels in between two rows of disks need to be smaller than those for pixels close to the disks.

The examples above are shown with Lindeberg's scheme. It should be mentioned that similar arguments hold for other unsupervised scale selection techniques [24,25,27–29,26,30–32], which select scales only from pre-specified image structures and do not take the underlying classification problems into account.

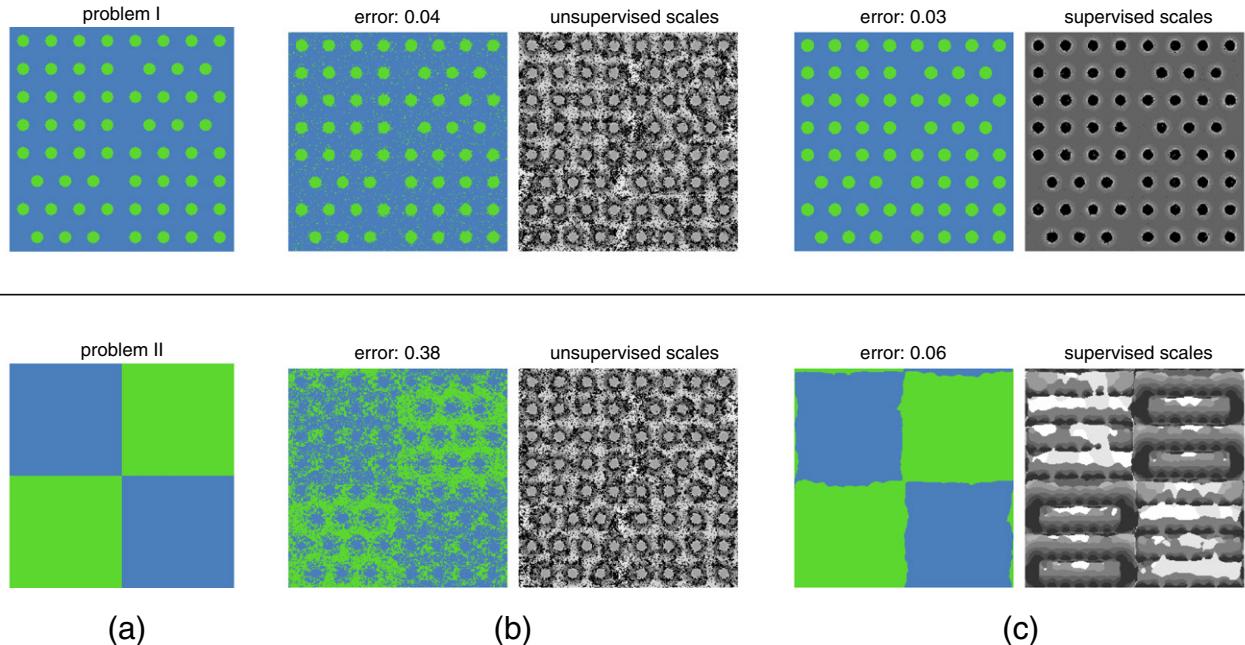


Fig. 3. Scale selection for two segmentation problems which have the same input image as in Fig. 2 but different labels. (a) Label images, which are the desired outputs defining two different segmentation tasks. (b) Classification results with Lindeberg's scale selection scheme. (c) Results with our method.

5. Max rule scale selection

We propose a procedure which incorporates scale selection into the classification problem itself. A parallel can still be drawn with Lindeberg's scheme, if we treat classifiers as a special type of image structure detectors.

5.1. Method

Since the right scales for a segmentation problem are not known a prior, we consider to conduct classification at all possible scales. Scales and the results accordingly at those scales are selected only after we have the classification at all scales. With this we aim to select supervised scales that are suitable for the segmentation problem.

The method of considering all scales follows a rationale in scale space, that is, if one cannot decide what scale should be included, then all the scales should be taken into consideration. Suppose we have n training images $\ell_j, j = 1, \dots, n$ and n label images $c_j, j = 1, \dots, n$, where $c_j(x, y) \in \{1, 2, \dots, K\}$ denotes the label at location (x, y) . At each scale σ , a classifier h_σ is trained with features extracted from the same scale. Specifically, features $F_j(x, y; \sigma)$ are extracted from training images ℓ_j represented at scale σ , and assigned the class label $c_j(x, y)$ at that pixel (x, y) . Note that at each pixel, the same label is assigned to features from all scales, $F_j(x, y; \sigma) = c_j(x, y), \forall \sigma$. The classifier, denoted as h_σ since it is dependent on scale σ , is trained as a mapping from the feature vectors $F_j(x, y; \sigma)$ to labels $c_j(x, y)$, making up a pair

$$(F_j(x, y; \sigma), c_j(x, y)), \quad (13)$$

for all x, y , and j . Methods of classifier training can be found in any standard pattern recognition textbook, for example, [58,57]. If QDA is used, then h_σ returns the class label of feature F as

$$h_\sigma(F) = \operatorname{argmax}_{k \in \{1-K\}} \delta_{\{k, \sigma\}}(F), \quad (14)$$

where $\delta_{\{k, \sigma\}}$ is the discriminant function of class ω_k at scale σ . As one classifier is obtained for each scale, in the end we obtain as many classifiers as the number of scales used.

Given an unseen image ℓ_t for testing, its labels at all locations need to be predicted. Since we do not know the appropriate scale for each pixel, we apply the trained classifiers to the test image at all scales and then select a scale for each pixel according to the so-called max rule. The max rule selects the scale at which the posterior probability attains the maximum across scale.

Specifically, at each scale σ , feature vectors $F_t(x, y; \sigma)$ are extracted for the test image ℓ_t . The trained classifier h_σ is then applied to the features and class labels are estimated as the outputs of the classifier

$$\hat{c}_t(x, y; \sigma) = h_\sigma(F_t(x, y; \sigma)). \quad (15)$$

A posterior probability $P(\hat{c}_t(x, y; \sigma) | F_t(x, y; \sigma))$ may be computed from the classification, either directly from the prior and class probability distributions as $P(\omega_k | F)$ in Eq. (8) for QDA, or more generally from other entities used in classification, such as discriminant functions or distances from the decision boundary [59]. The posterior can also be seen as a confidence measure of the classification result, and the larger it is, the more confident the classifier is.

Typically, different estimates $\hat{c}_t(x, y; \sigma)$ are obtained for different scales, since image structures change across scales. Ultimately, however, we are interested in one single overall classification, for which scale selection is necessary. We adopt the max rule for scale selection, which chooses the scale $S_t(x, y)$ at which the posterior achieves the maximum

$$S_t(x, y) = \operatorname{argmax}_{\sigma \in \mathbb{R}^+} P(\hat{c}_t(x, y; \sigma) | F_t(x, y; \sigma)). \quad (16)$$

Accordingly, the class label at the selected scale is assigned to the location (x, y) in the end

$$\hat{c}_t(x, y) = \hat{c}_t(x, y; \sigma = S_t(x, y)). \quad (17)$$

Our approach is schematically summarized in [Algorithm 1](#).

5.2. Reasons of using the max rule

The max rule is used since it fits well with the purpose of finding the scale with the maximum discrimination between the classes. Larger posteriors indicate clearer discrimination between classes and with the max rule (Eq. (16)), appropriate supervised scales can be selected. Moreover, the max rule is invariant to the parameterization of scale.

Furthermore, with the max rule, a parallel with Lindeberg's scheme can still be drawn. This can be done if we view classifiers as a special type of (nonlinear) filters, and posterior probabilities as filter responses at different scales (cf. Eqs. (7) and (16)).

6. Experimental investigations

The following methods are compared in the experiments: classification with unsupervised scales from Lindeberg's scheme, our method with the max rule scale selection, the best result from a single scale (*best scale*), and classification with concatenated features from all scales. The *best scale* is an optimistic estimate of the classification with features from a single scale. A classifier is trained for each scale and applied to a test image, and the result with the lowest error across scale is chosen. Since ground truth labels are not available for a test image and the best scales for different images can be different, it is not easy to identify the best scale in practice. Classification with concatenated features is commonly used to combine multi-scale information [60,61], which concatenates features extracted from various scales into a single feature vector and trains a classifier afterwards.

Algorithm 1. Supervised multi-scale image segmentation with the max rule scale selection.

Input: Training images $\ell_j(x, y)$ and labels $c_j(x, y), j = 1, \dots, n$, $c_j(x, y) \in \{1, \dots, K\}$; a sampling of the scales $\sigma_i \in \mathbb{R}^+, i = 1, \dots, s$; type of classifier (e.g., QDA or neural networks); a test image $\ell_t(x, y)$.

Output: Estimated labels $\hat{c}_t(x, y)$ and selected scales $S_t(x, y)$ for the test image $\ell_t(x, y)$.

Preprocessing: Compute the scale space representations of all images, i.e., $L_j(x, y; \sigma_i)$ and $L_t(x, y; \sigma_i)$, $\forall i, j$;

Training: Train a classifier h_{σ_i} at each scale;

for each $\sigma_i, i = 1, \dots, s$, **do**

- Extract features $F_j(x, y; \sigma_i)$, from all training images $L_j(x, y; \sigma_i), \forall j$;
- Train a classifier h_{σ_i} of the specified type from the pairs as in (13), the extracted features and the given labels.

Testing: Estimate labels $\hat{c}_t(x, y)$ and select scales $S_t(x, y)$;

for each $\sigma_i, i = 1, \dots, s$, **do**

- Extract features $F_t(x, y; \sigma_i)$ from the test image $L_t(x, y; \sigma_i)$;
- Apply the trained classifier h_{σ_i} to the extracted features $F_t(x, y; \sigma_i)$, and compute the class labels $\hat{c}_t(x, y; \sigma_i)$ with (15) and the associated posteriors $P(\hat{c}_t(x, y; \sigma_i) | F_t(x, y; \sigma_i))$.

From the classification at all scales, select the scales $S_t(x, y)$ with the max rule (16), at which the labels $\hat{c}_t(x, y)$ in (17) are used as the final class labels.

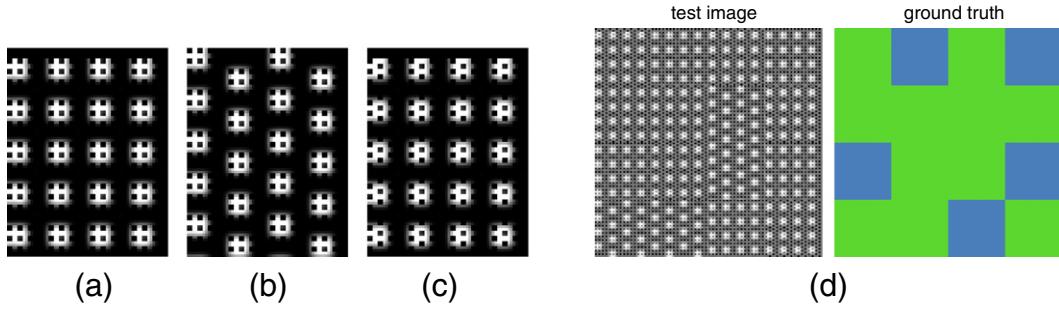


Fig. 4. Base textures and a test image. (a–c) Three base textures used to construct images. Texture a belongs to one class, and textures b and c to the other class. (d) An example image for testing and its ground truth label, where the two classes are represented by different colors.

Five data sets are used for evaluation in this section. The first is artificially constructed to illustrate the mechanism of our method, the second is to segment between four selected Brodatz textures [62], and the last three are real-world applications to segment rib and lungs in chest radiographs [43], or objects appearing as spots in live-cell fluorescence microscopy imaging [63,64], or vessels in retinal images [65,66].

For the first four data sets, QDA was used as the classifier and N -jet as features. We used standard QDA and N -jets in our experiment, but obviously our method is not limited to these choices. Twenty scales were sampled exponentially from the given scale range. Unsupervised scales were selected with the normalized Laplacian. For the data set of retinal vessel segmentation, the experimental setting in [66] was used, which will be explained in Subsection 6.5.

6.1. Multi-scale texture data

Using this artificial data, we intend to show the basic characteristics of the max rule scale selection for supervised segmentation. There are essentially two major scales in the data; at small or large scales, some structures can easily be segmented while others are difficult.

The data set consists of three different types of textures, shown in Fig. 4(a)–(c). The base textures in (a) and (b) are the same, but different from those in (c). The layout of the base textures are the same in (a) and (c), but different from those in (b). Thus it is difficult to discriminate (a) and (b) at small scales, and (a) and (c) at large scales. As a result, information from multiple scales is needed to segment all three textures.

Training and testing images were generated by putting square blocks of textures (a)–(c) together in a random way. Locations with texture (a) were assigned to one class, and the locations with texture (b) or (c) were assigned to the other class. An example image, which was from a test set, is shown in Fig. 4(d). Five images were used for training and the features were 6-jet. The scale range was [0.5, 50] for unsupervised scales and [0.5, 24] for both our method and the classification with concatenated features.

The classification results at individual scales are shown in Fig. 5. To save space, only 5 from 20 scales are displayed here. If we compare the results at scale 0.5 and scale 5.8, it can be observed that quite different classification results are obtained. That is because textures are discriminable at different scales, as analyzed before.

Fig. 6(a) gives the outcome of our method with the max rule scale selection. Compared with the results from individual scales, great improvement has been achieved. The selected supervised scales reflect the appropriate scales well to separate the textures. In comparison, the result with Lindeberg's scheme, shown in Fig. 6(b), is not very good. The unsupervised scales capture mainly the sizes of the base textures, which are not sufficient for this problem and thus lead to high error rate. Fig. 6(c) shows the classification with concatenated results, which fails to classify the textures corresponding to Fig. 4(b) correctly.

The first row of Table 1 shows the averaged error rates over 50 runs of experiments. The classification error on average can be reduced from 16.5% of the combined result with our max rule, while the error is 28.7% for unsupervised scales and 28.2% for concatenated features.

6.2. Brodatz texture segmentation

Multi-scale information is usually included to increase the discrimination ability of texture features [67,60]. In this experiment we show the segmentation of textures at various scales and the improved overall results by our scale selection approach.

We applied our method to images consisting of four types of Brodatz textures [62], namely D16, D20, D21, and D22, shown in Fig. 7(a). Every image was constructed as follows. First, generate a 2×2 matrix with elements randomly chosen from the set {1,2,3,4} and interpolate it to be of size 256×256 . Then at each pixel, sample from one of the four textures according to the pixel's gray-value. An example image with its ground truth is shown in Fig. 7(b). The features were 4-jet and ten images were used for training. The scale range was [0.5, 32] for our method

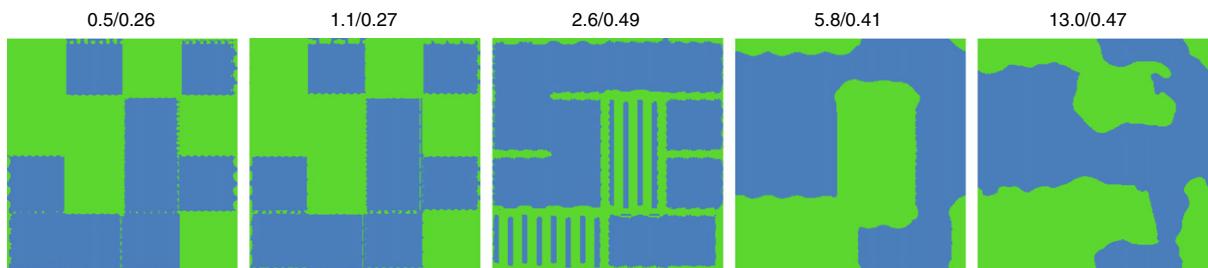


Fig. 5. Segmentation results by pixel classification at 5 different scales for the image in Fig. 4(d) from a test data set. The left number in the title of an image gives the feature scale σ_i and the right one gives the error rate for that image.

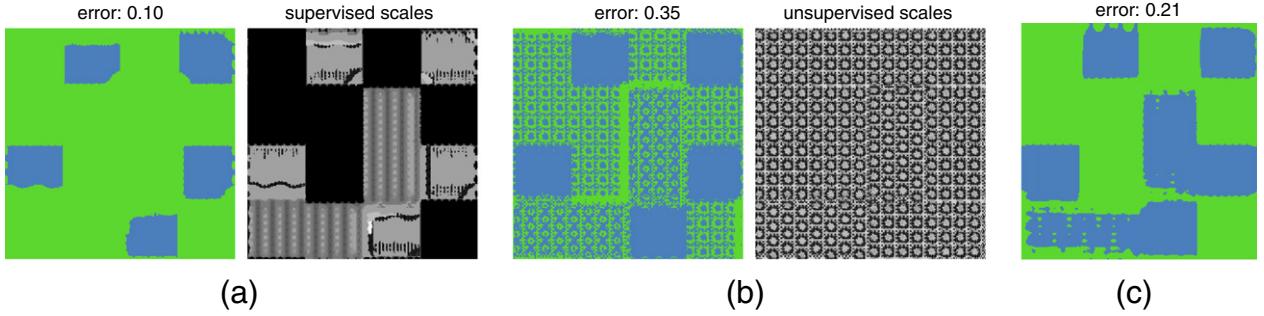


Fig. 6. Segmentation results and selected scales of (a) our method with the max rule, (b) the classification with Lindeberg's scale selection scheme. (c) Results with concatenated features from various scales.

Table 1

Error rates of different methods on four data sets, which are averaged over 50 trials for the first 3 data sets and 9 trials for the last one. “Unsupervised scale” denotes the classification with Lindeberg's scheme. “Scale concatenation” is the method which concatenates features from various scales. “Best scale” means that for each experiment, the result at the scale with the lowest error is selected. “Supervised scale” denotes our method, which selects scales with the max rule.

	Unsupervised scale	Scale concatenation	Best scale	Supervised scale
Multiscale-texture	28.7	28.2	16.5	10.3
Brodatz-texture	41.1	5.7	12.8	4.9
Rib&lung-seg	26.2	43.7	21.8	17.0
Spot-seg	11.1	11.9	5.8	7.7

and the classification with concatenated features, and [0.5, 50] for the method with Lindeberg's scheme.

Fig. 8 shows the classification results of the image in Fig. 7(b) at various scales. It can be observed that every texture is usually best discriminated from others in a certain scale range. For example, the texture at the upper-left corner (in yellow) is classified properly at

medium scales (e.g., scale 2.9), while the texture at the bottom-left corner (in blue) is classified properly at larger scales (e.g., scale 6.9), and the texture at the bottom-right corner (in red) at small scales (e.g., scale 1.2). Therefore, different scales are needed to separate these textures. This fact is clearly demonstrated by the scales selected by our max rule, shown in Fig. 9(a) on the right. The segmentation result with our method is displayed in Fig. 9(a) on the left, which considerably improves over the results of individual scales.

Fig. 9(b) shows the results with Lindeberg's scheme, whose selected scales reflect the blob sizes inside the textures. These scales, however, do not work well for texture segmentation. Fig. 9(c) shows the result of concatenating features from various scales. Its performance is quite comparable with our method, though it seems to segment the boundary pixels between different textures than our method less accurately.

The error rate comparison is shown in the second row of Table 1. The segmentation error on average can be reduced from 12.8% of the best individual scale to 4.9% of the combined result with the max rule, while the error with Lindeberg's scheme is 41.1%. The error for the classification with concatenated features is 5.7%, which is slightly larger than our method.

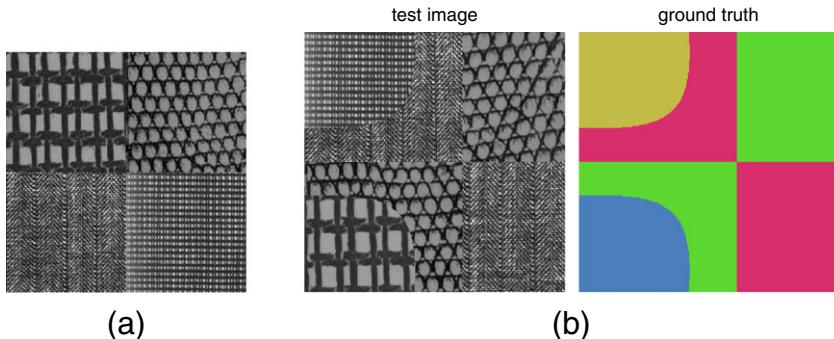


Fig. 7. Brodatz texture data. (a) Four types of Brodatz textures used in the experiments. (b) An example image from a test set and its ground truth label image.

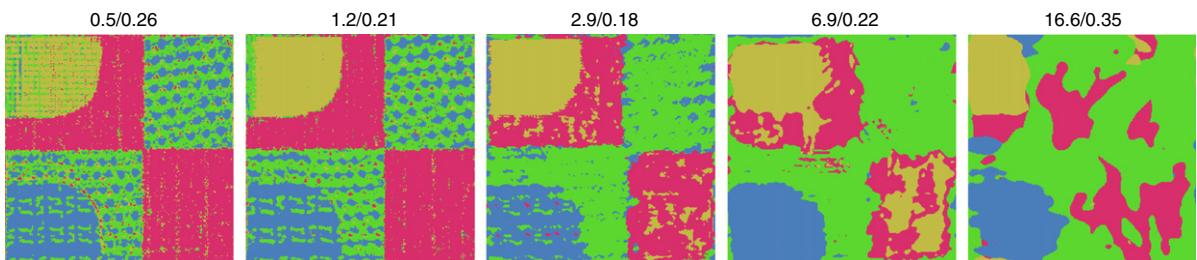


Fig. 8. Segmentation results by pixel classification at 5 different scales for the example image in Fig. 7(b) from a test data set. The left number in the title of an image gives the feature scale ϕ_i and the right one gives the error rate for that image.

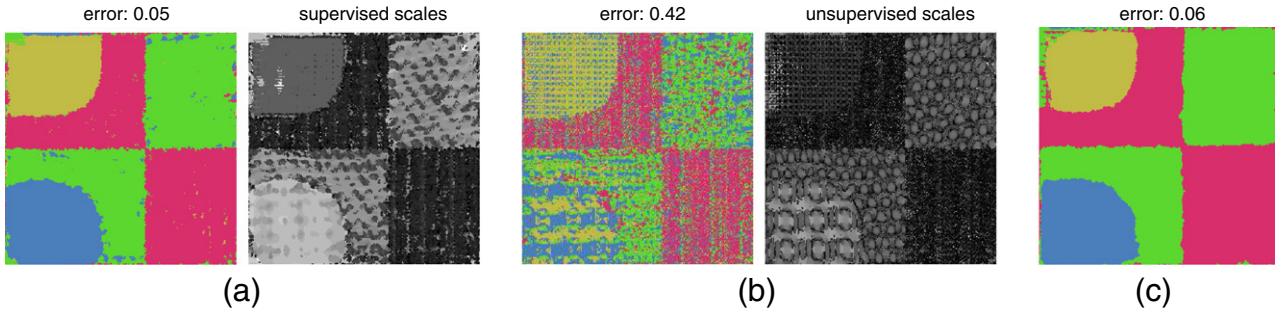


Fig. 9. Segmentation results and selected scales of (a) our method with the max rule, (b) the classification with Lindeberg's scale selection scheme. (c) Results with concatenated features from various scales.

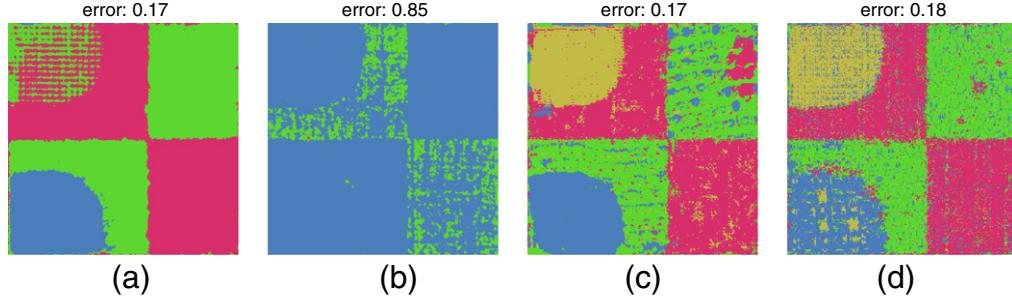


Fig. 10. Segmentation with different experimental settings. Classification with concatenated features by (a) changing scale range from [0.5,32] to [0.5,200], and (b) sampling 20 scales linearly from [0.5,32]. (c) Classification with our method when 2-jet was used as features. (d) Classification with Lindeberg's scheme when 8-jet was used as features.

In the following, we compare the results under different experiment settings: scale range and sampling, N -jets, and the number of training images.

Both the method with Lindeberg's scheme and our method were insensitive to scale ranges and sampling. Similar classification results were obtained when the scale range was changed to [0.5,200]. The same conclusion held when 20 scales were sampled linearly, instead of exponentially. In comparison, the method with concatenated features was heavily influenced by the scale parameterization. When the scale range was changed to [0.5,200], the segmentation results became worse (Fig. 10(a)). When scales were sampled linearly, instead of exponentially, within the same range [0.5,32], the segmentation results deteriorated dramatically (Fig. 10(b)).

When lower-order jets (e.g., 2-jet) were used as features instead of 4-jet, the classification results became more noisy and less accurate for all the classification methods. As an example, the result of our method is shown in Fig. 10(c). When higher-order jets such as 6-jet or 8-jet were used, our method improved slightly upon Fig. 9(a): the segmentation became less noisy inside the region of the same texture. When 6-jet was used, the method with concatenated features classified all pixels into one class, possibly due to the high-dimensionality of the features. When higher-order jets were used, the method with Lindeberg's scheme improved its performance considerably. The result with 8-jet is shown in Fig. 10(d), which reduces the error rate by a half compared to Fig. 9(b), but is still inferior than the method with concatenated features and our method. Using 10-jet or 14-jet did not improve further upon 8-jet.

For all three methods, increasing the training images from 10 to 20 or 50 did not lead to significant improvements. When 5 images were used, the method with concatenated features and our method deteriorated about two percent in error rate.

6.3. Rib and lung segmentation

We use the data set from [43], and the task is to segment ribs and lungs from the background. See Fig. 11 for an example. In total there

were 30 images, and in our experiment 25 randomly chosen ones were used in training and the rest 5 in testing. The features were 6-jet, and the scale range was [0.5,32] for our method and the method with concatenated features, and [0.5,200] for the method with Lindeberg's scheme.

Fig. 12 shows the segmentation at individual scales of the image in Fig. 11, which was from a test set. Ribs and lungs are best discriminated at medium scales (e.g., scales 2.9 and 6.9), and the background is best segmented at large scales (e.g., 32). This is also reflected in the supervised scales selected by our max rule, see Fig. 13(a) on the right. Improved classification is obtained with the supervised scales, as shown in Fig. 13(a) on the left.

Fig. 13(b) shows the result with Lindeberg's scheme. It seems that the scale selector based on the Laplacian cannot capture the structures inside the lung areas, which are instead treated as two large blobs. As a result, the discrimination between the ribs and lungs is lost and the results are not satisfactory.

Concatenating features segments relatively well between the ribs and lungs, but not very well for the boundary areas between the lung areas and the background, as shown in Fig. 13(c). Thus it returns a high overall error rate.

On average, the error rate with the max rule scale selection is 17.0%, which improves from 21.8% of the best individual and 26.2%

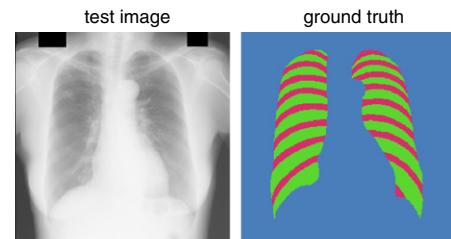


Fig. 11. Example input and output images from the rib and lung segmentation data (the 30th image), which is a three-class problem.

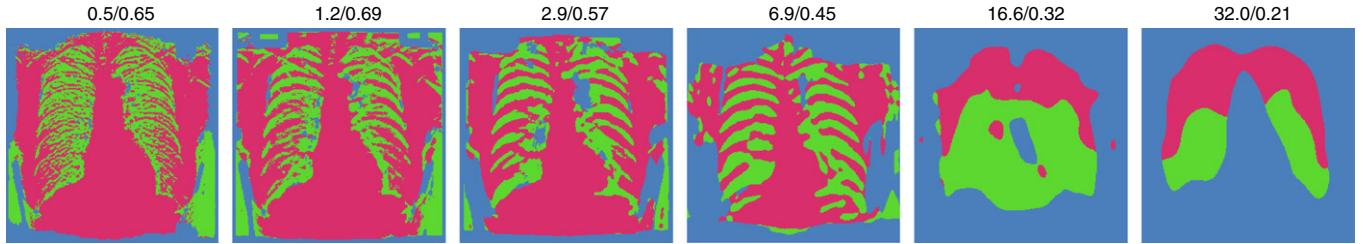


Fig. 12. Segmentation results by pixel classification at 6 different scales for the example image in Fig. 11 from a test data set. The left number in the title of an image gives the feature scale α_i and the right one gives the error rate for that image.

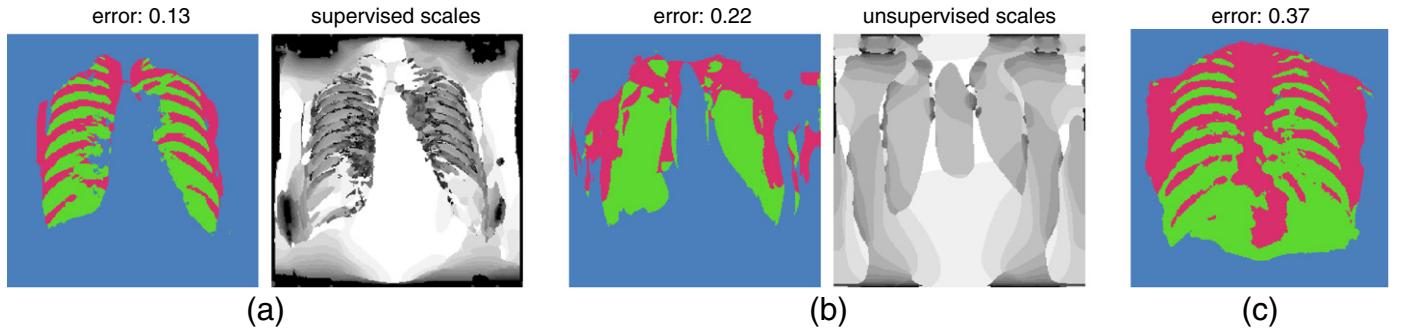


Fig. 13. Segmentation results and selected scales of (a) our method with the max rule, (b) the classification with Lindeberg's scale selection scheme. (c) Results with concatenated features from various scales.

of the method with Lindeberg's scheme, as shown in the third row of Table 1. The error rate for the concatenated features is 43.7%.

By changing the scale range from [0.5,200] to [0.5,50], the method with Lindeberg's scheme became slightly worse, as large structures such as the lung areas may need scales larger than 50. Our method did not change much after changing its scale range to [0.5,200]. In comparison, the method with concatenated features was heavily influenced by the scale parameterization, as shown in Fig. 14. By changing the scale range from [0.5,32] to [0.5,200], it segmented the lung areas better from the background, but less accurately between the ribs and lungs (Fig. 14(a)). Sampling scales linearly (instead of exponentially) from the original range [0.5,32], which implicitly increased the influence of large scales, led to even better segmentation between the lung areas and the background (Fig. 14(b)). We also considered changing both scale range and sampling, that is, sampling scales linearly from the range [0.5,200]. In this new setting the ribs and lungs could no longer be discriminated, as features at very large scales (e.g., close to 200) were not very informative about the small structures (Fig. 14(c)).

For both the method with Lindeberg's scheme and our methods, increasing the order of N -jet to 8 or 10 improved the results slightly,

but not significantly. Our method had comparable results when the order was 2 or 4, but deteriorated heavily when 1-jet was used. The method with unsupervised scales became much worse when 2-jet was used, where the whole image was classified into background. The results with concatenated features did not change much when 2-jet or 4-jet was used, but became bad for 1-jet. When 8-jet was used, the results also became much worse, possibly due to the high dimensionality of the features and the limited training samples.

When the number of training images decreased from 25 to 5, the classification deteriorated only a little for all three methods, a few percentage in error rates. However, when the number decreased to 1 or 2, the segmentation results became much worse and typically most of the rib and lung areas were classified into the background.

6.4. Spot segmentation

Spot detection is an important problem in live-cell fluorescence microscopy imaging, where many subresolution objects appear as diffraction-limited spots [63,64]. It is a challenging problem because images have low signal-to-ratio and low resolution. In the experiments

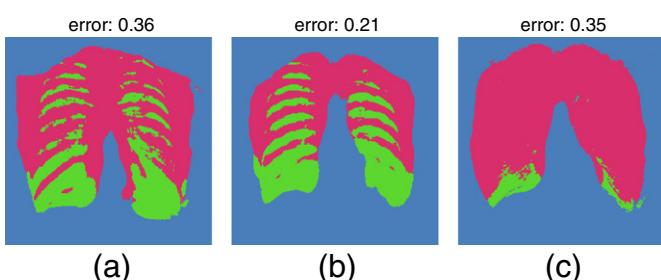


Fig. 14. Segmentation with different experimental settings. Classification with concatenated features by (a) changing scale range from [0.5,32] to [0.5,200], (b) sampling 20 scales linearly from [0.5,32], and (c) sampling 20 scales linearly from a larger range [0.5,200].

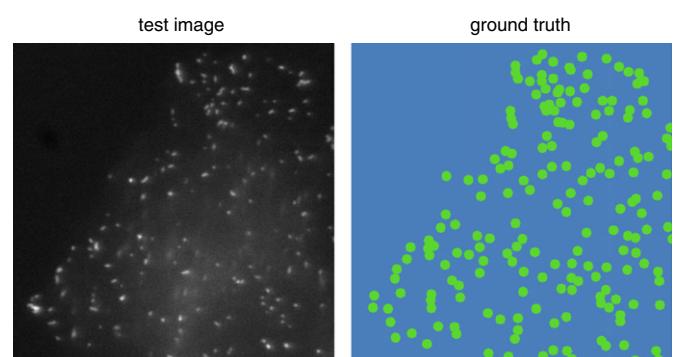


Fig. 15. An image from the spot segmentation problem.

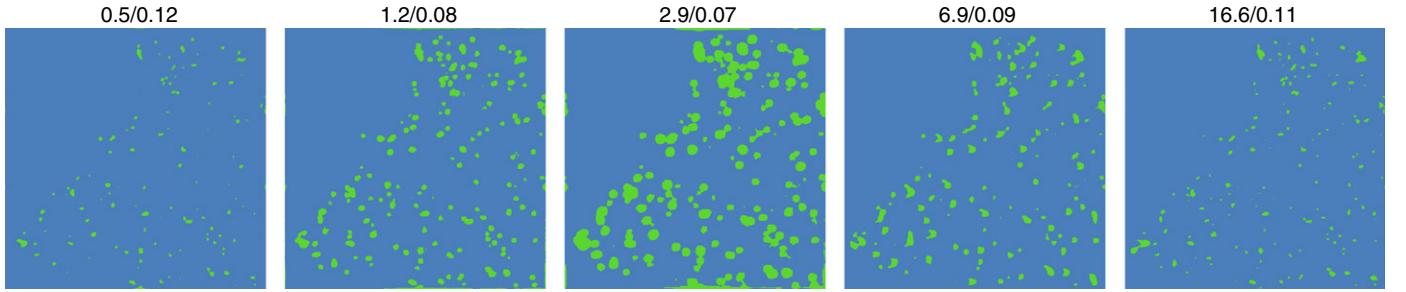


Fig. 16. Segmentation results by pixel classification at 5 different scales for the example image in Fig. 15 from a test data set. The left number in the title of an image gives the feature ϕ_i and the right one gives the error rate for that image.

nine images were used, labeled by experts with a dot representing each spot. We approximated the spot with a blob centered in the annotated dot, and the task was to segment the spots from the background. One example image is shown in Fig. 15. Each experiment used 1 image for training and the rest 8 images for testing. The features were 2-jet and scale range was [0.5, 32] for all different approaches.

Fig. 16 shows the segmentation at individual scales of the image in Fig. 15 from a test set. We can see that the results vary a lot across the scales. The combined result with our max rule scale selection is shown in Fig. 17(a), whose error rate is comparable with the best

segmentation from individual scales. Small scales are selected for the pixels in the center of the spots, while relatively large scales for the surrounding pixels. This is because of the high noise level in the image: the center of spots are quite bright and thus segmented accurately at small scales, while the surrounding areas need more blurring to identify the class labels. Similar supervised scales have been selected in Fig. 1(c), the last row with highest noise level.

Fig. 17(b) shows the results with unsupervised scales. It selects quite different scales than those in Fig. 17(a) and its error rate is significantly higher. Feature concatenation leads to reasonably good results as

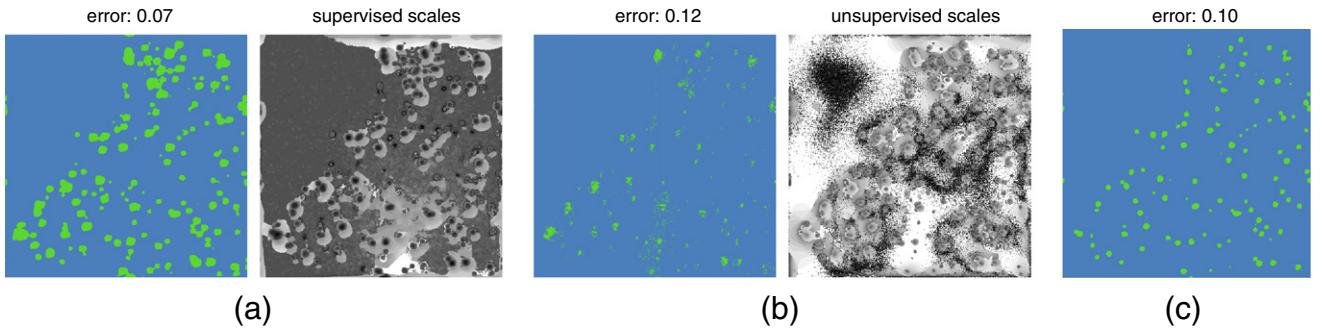


Fig. 17. Segmentation results and selected scales of (a) our method with the max rule, (b) the classification with Lindeberg's scale selection scheme. (c) Results with concatenated features from various scales.

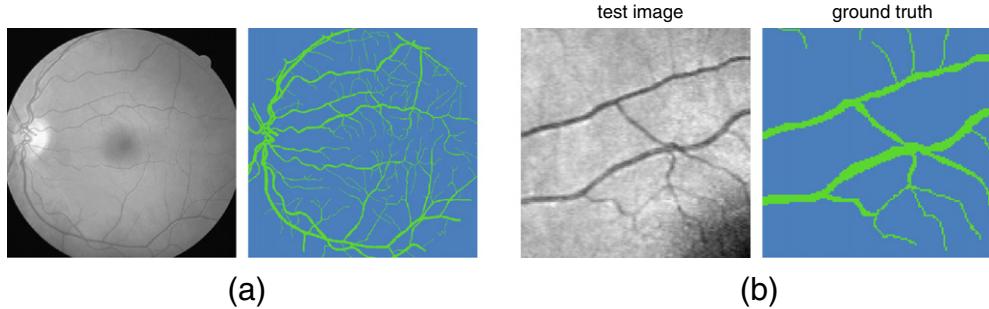


Fig. 18. Retinal vessel segmentation data set. (a) A whole image and its ground truth labels. (b) An enlarged part of a.

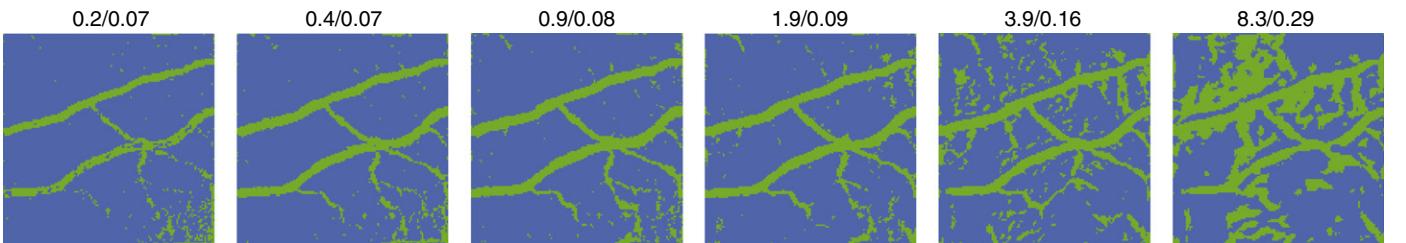


Fig. 19. Segmentation results by pixel classification at 6 different scales for the example image in Fig. 18(b) from a test data set. The left number in the title of an image gives the feature scale ϕ_i and the right one gives the error rate for that image.

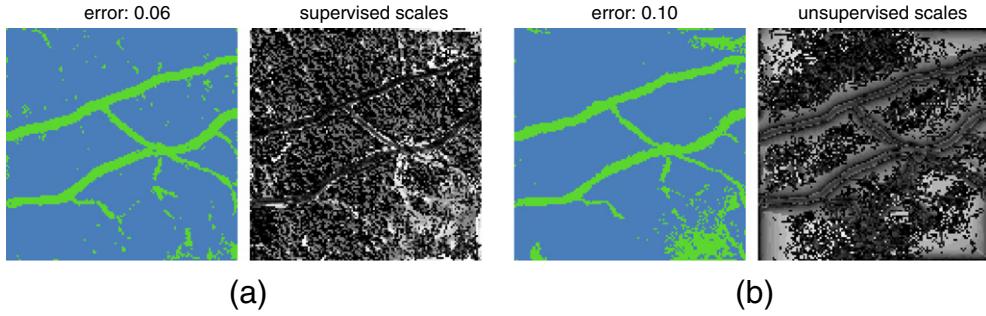


Fig. 20. Segmentation results and selected scales of (a) our method with the max rule, (b) the classification with Lindeberg's scale selection scheme.

shown in Fig. 17(c), with slightly higher error rates. Repeating the experiments by using each of the 9 images for training and the rest 8 for testing, the average error rates are 5.8% for the best results from individual scales, 7.7% for our method, 11.1% for the method with unsupervised scales, and 11.9% for the scale concatenation, as shown in Table 1.

6.5. Retinal vessel segmentation

Segmenting retinal vessels from the background is important for eye disease detection and diagnosis. We used the DRIVE (Digital Retinal Images for Vessel Extraction) data set from the Image Sciences Institute, Utrecht University, The Netherlands [65,66]. There are 40 images in total with manual annotation, and in each experiments, we randomly choose 5 images for training and another 20 for testing. One example image is shown in Fig. 18(a).

The same classification setting as in [66] was used. Compared to experiments in previous subsections, three things are different here. First, the classifier is the 30-nearest neighbor method. Second, unsupervised scales for the retinal vessels are selected with a ridge detector [19], whose scale-normalized form (cf. Eq. (5)) is $\sigma^{8\gamma}(L_{xx} + L_{yy})^2 \left((L_{xx} - L_{yy})^2 + L_{xy}^2 \right)$ with $\gamma = \frac{3}{4}$. Third, the basic features are already the concatenation of 2-jet from five scales $\{\frac{1}{4}, \frac{1}{2}, 1, 2, 4\}$ and the pixel gray value. If a scale s is selected for a pixel, then the concatenation of 2-jet from scales $\{\frac{s}{4}, \frac{s}{2}, s, 2s, 4s\}$ is used as features. The scale range for our method was [0.2, 12], while that for unsupervised scales was [0.5, 50].

Fig. 19 shows the segmentation results of the image in Fig. 18(b) at different scales. The results differ a lot across the scales; at small scales, the segmented vessels are a little noisy and not connected well, while at large scales, many background pixels are misclassified as vessels. Combining the results with the max rule leads to good segmentation of the vessels, as shown in Fig. 20(a). Small scales are selected for the center pixels of the vessels as they can be separated from the background most accurately without much blurring, while relatively large scales are needed for the surrounding pixels. The method with unsupervised scale selection leads to reasonably good results, though quite some background pixels are misclassified as vessels. Its selected scales reflect the characteristic sizes of the ridges. Repeating the experiments for 50 times, the average error rates are 10.8% for the best results from individual scales, 10.6% for our method with the max rule, and 17.1% for the method with unsupervised scales.

7. Discussions

The spatial information in our segmentation model is implicitly incorporated in the feature N -jet, which effectively smooths the image before computing the derivatives. As a result, spatially close locations tend to return similar features and thus similar classification results. See for example the classification results at individual scales in Figs. 5, 8, and 12. At scale 0.5, images are blurred only a little bit and thus the segmentations look quite spiky. At increasing scales,

segmentation results become more smoothed as images are blurred more heavily.

The segmentation of our results is still a little noisy as the max rule, based on the classification from different scales, selects a scale for each pixel independently. See Figs. 9(a) and 13(a) as examples. A smoothing constraint may be explicitly applied to the segmentation to improve the results. One possibility is to embed our method into a conditional random field framework [68,1–3], which models the spatial constraints of pixels in an image.

From the results, see for example Figs. 9(a) and 13(a), it seems that the edges are quite sharp and preserved very well. The reason may lie in the fact that small scales are usually selected for the boundary pixels and thus their segmentation results are based only on a little blurring. Take the supervised scales for the rib and lung segmentation as an example (Fig. 13(a) on the right). Though medium scales are selected for the rib and lung areas, smaller scales are selected for their boundary pixels. This property may be connected to *feature localization with automatic scale selection* [18], which selects a scale to localize an image structure which is previously detected at another (especially coarse) scale. Typically a smaller scale is selected for localization.

The max rule relates our supervised approach in a nice way to Lindeberg's classical scheme, but other so-called classifier combining rules may certainly be considered as well [69,70]. Some care should be taken, however. The max rule has the invariant property of scale parameterization. Rules that do not have this invariant property may display large variations in performance depending on the particular scale parameterization and scale range. How to make appropriate choices for these or how to adapt known combining schemes properly to the situation considered in this work remain for further study.

It should be mentioned that although the max rule is used to select scales for supervised image segmentation in this paper, it may be applied to other applications as well. For example, in image classification, we may extract features from various scales and train a classifier for each scale. In the test phase, an image is classified at all scales and the result at the scale with the maximum posterior is chosen as the final classification.

8. Conclusion

This paper deals with the problem of selecting proper scales for supervised image segmentation problems. We propose to adapt the selected scales to the problem under consideration in a supervised way and to use the outputs from the classifier to select the right scales for the classification problem. The experiments show that our method can indeed select appropriate scales for supervised segmentation, and improve upon the results employing individual scales, unsupervised scales from Lindeberg's scale selection scheme, and concatenated features from all scales.

Acknowledgments

The authors would like to thank the reviewers for their constructive comments. Ihor Smal, Erik Meijering (both Erasmus Medical

Centre) and Ilya Grigoriev and Anna Akhmanova (both Utrecht University) are kindly acknowledged for the spot detection data set. Marco Loog was partly supported by the Research Grant Program of the Faculty of Science, University of Copenhagen (research grant 10-05087) and the Innovative Research Incentives Scheme of The Netherlands Research Organization (NWO, VENI grant 639.021.611).

References

- [1] S. Kumar, M. Hebert, Discriminative random fields: a discriminative framework for contextual interaction in classification, in: IEEE Conf. Computer Vision, 2003, pp. 1150–1157.
- [2] X. He, R. Zemel, M. Carreira-Perpinán, Multiscale conditional random fields for image labeling, in: IEEE Conf. Computer Vision and Pattern Recognition, 2004, pp. 695–702.
- [3] P. Krahenbuhl, V. Koltun, Efficient inference in fully connected CRFs with Gaussian edge potentials, in: Neural Information Processing Systems, 2011.
- [4] J. Shotton, J. Winn, C. Rother, A. Criminisi, TextronBoost for image understanding: multi-class object recognition and segmentation by jointly modeling texture, layout, and context, Int. J. Comput. Vis. 81 (1) (2009) 2–23.
- [5] M. García, D. Puig, Supervised texture classification by integration of multiple texture methods and evaluation windows, Image Vis. Comput. 25 (7) (2007) 1091–1106.
- [6] M. Loog, Supervised dimensionality reduction and contextual pattern recognition in medical image processing, Ph.D. Thesis, Utrecht University, 2004.
- [7] N. Vandebroucke, L. Macaire, J. Postaire, Color image segmentation by pixel classification in an adapted hybrid color space, Comp. Vision Image Underst. 90 (2) (2003) 190–216.
- [8] D. Geman, S. Geman, C. Graffigne, P. Dong, Boundary detection by constrained optimization, IEEE Trans. Pattern Anal. Mach. Intell. 12 (7) (1990) 609–628.
- [9] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (8) (2000) 888–905.
- [10] A. Witkin, Scale-space filtering: a new approach to multi-scale description, in: Int. Joint Conf. Artificial Intelligence, Vol. 9, 1983, pp. 1019–1023.
- [11] J. Koenderink, The structure of images, Biol. Cybern. 50 (5) (1984) 363–370.
- [12] T. Lindeberg, Scale-Space Theory in Computer Vision, Kluwer Academic, 1994.
- [13] B. Ter Haar Romeny, Front-End Vision and Multi-Scale Image Analysis, Kluwer Academic, 2002.
- [14] P. Burt, E. Adelson, The Laplacian pyramid as a compact image code, IEEE Trans. Commun. 31 (4) (1983) 532–540.
- [15] S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (7) (1989) 674–693.
- [16] L. Florack, The Syntactical Structure of Scalar Images, Universiteit Utrecht, 1993.
- [17] F. Bergholm, Edge focusing, IEEE Trans. Pattern Anal. Mach. Intell. 9 (6) (1987) 726–741.
- [18] T. Lindeberg, Feature detection with automatic scale selection, Int. J. Comput. Vis. 30 (2) (1998) 79–116.
- [19] T. Lindeberg, Edge detection and ridge detection with automatic scale selection, Int. J. Comput. Vis. 30 (2) (1998) 117–154.
- [20] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
- [21] K. Mikolajczyk, C. Schmid, Scale & affine invariant interest point detectors, Int. J. Comput. Vis. 60 (1) (2004) 63–86.
- [22] O. Chomat, V.C. de Verdière, D. Hall, J.L. Crowley, Local scale selection for Gaussian based description techniques, in: 6th European Conf. Computer Vision, 2000, pp. 117–133.
- [23] R. Collins, Mean-shift blob tracking through scale space, in: IEEE Conf. Computer Vision and Pattern Recognition, 2003, pp. 234–240.
- [24] T. Kadir, M. Brady, Saliency, scale and image description, Int. J. Comput. Vis. 45 (2) (2001) 83–105.
- [25] K. Okada, D. Comaniciu, A. Krishnan, Scale selection for anisotropic scale-space: application to volumetric tumor characterization, in: IEEE Conf. Computer Vision and Pattern Recognition, 2004, pp. 594–601.
- [26] M. Loog, K. Pedersen, B. Markussen, Maximum likely scale estimation, in: Deep Structure, Singularities, and Computer Vision, LNCS, 3753, 2005, pp. 146–156.
- [27] D. Strong, J. Aujoil, T. Chan, et al., Scale recognition, regularization parameter selection, and Meyer's G norm in total variation regularization, Multiscale Model Sim. 5 (1) (2007) 273–303.
- [28] B. Luo, J. Aujoil, Y. Gousseau, Local scale measure from the topographic map and application to remote sensing images, Multiscale Model Sim. 8 (1) (2009) 1–29.
- [29] J. Sporrings, J. Weickert, Information measures in scale-spaces, IEEE Trans. Inf. Theory 45 (3) (1999) 1051–1058.
- [30] T. Brox, J. Weickert, A TV flow based local scale measure for texture discrimination, 8th European Conf. Comput. Vis. (2004) 578–590.
- [31] B. Hong, S. Soatto, K. Ni, T. Chan, The scale of a texture and its application to segmentation, in: IEEE Conf. Computer Vision and Pattern Recognition, 2008, pp. 1–8.
- [32] S. Boltz, F. Nielsen, S. Soatto, Texture regimes for entropy-based multiscale image analysis, in: European Conf. Computer Vision, 2010, pp. 692–705.
- [33] Y. Wu, C. Guo, S. Zhu, From information scaling of natural images to regimes of statistical models, Quart. Appl. Math. 66 (2008) 81–122.
- [34] C. Guo, S. Zhu, Y. Wu, Primal sketch: integrating structure and texture, Comp. Vision Image Underst. 106 (1) (2007) 5–19.
- [35] K. Suzuki, H. Abe, H. MacMahon, Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network, IEEE Trans. Med. Imaging 25 (4) (2006) 406–416.
- [36] M. Loog, B. van Ginneken, A. Schilham, Filter learning: application to suppression of bony structures from chest radiographs, Med. Image Anal. 10 (6) (2006) 826–840.
- [37] M. Loog, Y. Li, D. Tax, Maximum membership scale selection, in: 8th Int'l Workshop on Multiple Classifier Systems, LNCS, 5519, 2009, pp. 468–477.
- [38] Y. Li, D. Tax, M. Loog, Supervised scale-invariant segmentation (and detection), in: Scale Space and Variational Methods in Computer Vision, 2012, pp. 350–361.
- [39] J. Koenderink, A. van Doorn, Receptive field families, Biol. Cybern. 63 (4) (1990) 291–297.
- [40] L. Florack, B. ter Haar Romeny, J. Koenderink, M. Viergever, Scale and the differential structure of images, Image Vis. Comput. 10 (6) (1992) 376–388.
- [41] J. Folkesson, E. Dam, O. Olsen, P. Pettersen, C. Christiansen, Segmenting articular cartilage automatically using a voxel classification approach, IEEE Trans. Med. Imaging 26 (1) (2007) 106–115.
- [42] B. Van Ginneken, M. Stegmann, M. Loog, Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database, Med. Image Anal. 10 (1) (2006) 19–40.
- [43] M. Loog, B. Ginneken, Segmentation of the posterior ribs in chest radiographs using iterated contextual pixel classification, IEEE Trans. Med. Imaging 25 (5) (2006) 602–611.
- [44] J. Koenderink, A. Van Doorn, Representation of local geometry in the visual system, Biol. Cybern. 55 (6) (1987) 367–375.
- [45] L. Florack, B. ter Haar Romeny, J. Koenderink, M. Viergever, Cartesian differential invariants in scale-space, J. Math. Imaging Vis. 3 (4) (1993) 327–348.
- [46] C. Schmid, R. Mohr, Local grayvalue invariants for image retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 19 (5) (1997) 530–535.
- [47] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, Speeded-up robust features (SURF), Comput. Vis. Image Underst. 110 (3) (2008) 346–359.
- [48] S. Lazebnik, C. Schmid, J. Ponce, A sparse texture representation using local affine regions, IEEE Trans. on Pattern Anal. Mach. Intell. 27 (8) (2005) 1265–1278.
- [49] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Trans. Pattern Anal. Mach. Intell. 27 (10) (2005) 1615–1630.
- [50] R. Fergus, P. Perona, A. Zisserman, Object class recognition by unsupervised scale-invariant learning, in: IEEE Conf. Computer Vision and Pattern Recognition, 2003, pp. 264–271.
- [51] R. Manmatha, J. Rothfeder, A scale space approach for automatically segmenting words from historical handwritten documents, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1212–1225.
- [52] J. Maver, Self-similarity and points of interest, IEEE Trans. Pattern Anal. Mach. Intell. 32 (7) (2010) 1211–1226.
- [53] D. Panda, A. Rosenfeld, Image segmentation by pixel classification in (gray level, edge value) space, IEEE Trans. Comput. 100 (9) (1978) 875–879.
- [54] W. Blanz, E. Reinhardt, Image segmentation by pixel classification, Pattern Recogn. 13 (4) (1981) 293–298.
- [55] S. Haring, M. Viergever, J. Kok, Kohonen networks for multiscale image segmentation, Image Vis. Comput. 12 (6) (1994) 339–344.
- [56] Z. Yu, H. Wong, G. Wen, A modified support vector machine and its application to image segmentation, Image Vis. Comput. 29 (1) (2011) 29–40.
- [57] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley-Interscience Publication, 2000.
- [58] C. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
- [59] R. Duin, D. Tax, Classifier conditional posterior probabilities, in: Proc. Joint IAPR Int. Workshops SSPR'98 and SPR'98, LNCS, 1451, 1998, pp. 611–619.
- [60] J. Mao, A. Jain, Texture classification and segmentation using multiresolution simultaneous autoregressive models, Pattern Recogn. 25 (2) (1992) 173–188.
- [61] S. Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, in: IEEE Conf. Computer Vision and Pattern Recognition, 2006, pp. 2169–2178.
- [62] P. Brodatz, Textures: A Photographic Album for Artists & Designers, Dover, New York, 1966.
- [63] I. Smal, M. Loog, W. Niessen, E. Meijering, Quantitative comparison of spot detection methods in five-cell fluorescence microscopy imaging, in: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE, 2009, pp. 1178–1181.
- [64] I. Smal, M. Loog, W. Niessen, E. Meijering, Quantitative comparison of spot detection methods in fluorescence microscopy, IEEE Trans. Med. Imaging 29 (2) (2010) 282–301.
- [65] J. Staal, M. Abràmoff, M. Niemeijer, M. Viergever, B. Van Ginneken, Ridge-based vessel segmentation in color images of the retina, IEEE Trans. Med. Imaging 23 (4) (2004) 501–509.
- [66] M. Niemeijer, J. Staal, B. Van Ginneken, M. Loog, M. Abramoff, Comparative study of retinal vessel segmentation methods on a new publicly available database, in: Proceedings of SPIE Medical Imaging, Vol. 5370, 2004, pp. 648–656.
- [67] H. Greenspan, R. Goodman, C.H. Anderson, Learning texture discrimination rules in a multiresolution system, IEEE Trans. Pattern Anal. Mach. Intell. 16 (9) (1994) 894–901.
- [68] J.D. Lafferty, A. McCallum, F.C.N. Pereira, Conditional random fields: probabilistic models for segmenting and labeling sequence data, in: 18th Int. Conf. Machine Learning, 2001, pp. 282–289.
- [69] J. Kittler, M. Hatef, R. Duin, J. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239.
- [70] L. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms, Wiley-Interscience, 2004.