A Multiscale Approach to Image Segmentation Using Kohonen Networks

S. Haring, M.A. Viergever, and J.N. Kok

RUU-CS-93-06
February 1993

Utrecht University

Department of Computer Science

Padualaan 14, P.O. Box 80.089,
3508 TB Utrecht, The Netherlands,
Tel. : ... + 31 - 30 - 531454

A Multiscale Approach to Image Segmentation Using Kohonen Networks

S. Haring, M.A. Viergever, and J.N. Kok

Technical Report RUU-CS-93-06
February 1993

Department of Computer Science
Utrecht University
P.O.Box 80.089
3508 TB Utrecht
The Netherlands

ISSN: 0924-3275

A Multiscale Approach to Image Segmentation Using Kohonen Networks

S. Haring[1,2] M.A. Viergever[2] J.N. Kok[1]

[1] Department of Computer Science, Utrecht University,
P.O.Box 80.089, 3508 TB Utrecht, The Netherlands, E-mail bas@cs.ruu.nl

[2] Computer Vision Research Group, Utrecht University Hospital,
Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

Abstract

An approach is developed to multiscale image segmentation, based on pixel classification by means of a Kohonen network. An image is described by assigning a feature pattern to each pixel, consisting of a scaled family of differential geometrical invariant features. The invariant feature pattern representation of a training image is input to a Kohonen network, in order to obtain a description of the feature space in terms of so-called *prototypical feature patterns* (the weight vectors of the network). Supervised labeling of these prototypical feature patterns may be accomplished using classes derived from an a priori segmentation of the training image. We can segment any image similar to the training image by comparing the feature pattern representation of each pixel with all weight vectors, and assigning each pixel the class of the best matching weight vector. In our study we evaluated the benefit of applying features at multiple scales, as well as the effects of first- and second order information on the results.

1 Introduction

Many image processing applications have the aim of extracting information from image data. Image segmentation is a crucial step in these applications and can be defined as dividing an image into several parts which are intended to refer to one object. The quality of both volumetric visualization and quantitative analyses strongly depends on the segmentation of the object to be displayed or measured.

This paper describes a neural network based approach to image segmentation, by which the segments are acquired by pixel classification. A principle idea of our approach is that, while local image structure is known to be represented completely by a family of scaled differential operators [KvD84, KvD90, FtHRKV92, tHRFKV91], there is hardly any knowledge about which geometric invariants at which scale are optimal or even suitable for specific image processing tasks as, e.g. image segmentation.

Each pixel of the image to be analyzed will be assigned a feature pattern consisting of a limited number of differential geometrical invariants at various scales. We assume that a segment of the image corresponds to a separable and coherent region in the feature space [BC91]. We are concerned with ascertaining these regions, and using them as a basis for classification. Since the regions are not necessarily linearly separable we might not be

able to delineate them using conventional pattern recognition methods. For this reason we applied a supervised classification strategy based on *Kohonen networks* [Koh84].

The objective of the study is threefold: (i) to investigate the benefits of the multiscale approach; (ii) to analyze the effects of various differential geometrical invariants (up to the second order) on the classifications; (iii) to evaluate whether a Kohonen network is an appropriate tool for image segmentation.

2 Image description

Scale plays an important role in image understanding. An image, being a physical observable of a scene, represents the scene on a finite range of scales only. The sampling characteristics determine the lower bound (inner scale), the scope of the field of view the upper bound (outer scale). The aim of image understanding is to infer scene structure; this calls for a multiscale description of image structure. It has been shown [FtHRKV92] that the only operator family satisfying the natural front-end vision constraints of linearity, shift-, rotation-, and scale invariance is the Gaussian and all its partial derivatives; this operator family provides a complete representation of image structure.

The zeroth order operator of the Gaussian differential operator family is the well known *Gaussian* function

$$G_\sigma(\vec{x}) = \frac{1}{(\sigma\sqrt{2\pi})^D} e^{-\frac{(\vec{x}\cdot\vec{x})}{2\sigma^2}}$$

where D is the spatial dimension of the image; $\vec{x} = (x_1, \ldots, x_D)$, and σ is the width of the Gaussian. Assume an image with luminance pattern $L(\vec{x})$. The property

$$\frac{\partial L(\vec{x})}{\partial x_i} \otimes G_\sigma(\vec{x}) = \frac{\partial G_\sigma(\vec{x})}{\partial x_i} \otimes L(\vec{x})$$

shows that all image derivatives can be obtained by convolving with the corresponding derivative of the Gaussian. Consequently, a family of scaled differential operators is available, by which image structures may be described completely up to any desired order at various scales.

We are interested in image features which are invariant under certain transformations. We confine ourselves to features which are invariant under orthogonal co-ordinate transformations, i.e. translation and rotation. An infinite number of these invariants can be constructed, but it is a straightforward, although not trivial, matter to derive so-called *irreducible invariants* on which all others depend [FtHRKV92]. For two-dimensional images there are five irreducible invariants, which may be represented in various ways. An often used representation is the tensorial notation,

$$\{L, L_i L_i, L_{ii}, L_i L_{ij} L_j, L_{ij} L_{ji}\}$$

or in Cartesian notation:

$$\left\{L, L_x^2 + L_y^2, L_{xx} + L_{yy}, L_x^2 L_{xx} + L_y^2 L_{yy} + 2L_x L_y L_{xy}, L_{xx}^2 + L_{yy}^2 + 2L_{xy}^2\right\}$$

Note that L is just the image intensity (or luminance), $L_i L_i$ is the squared norm of the gradient, and L_{ii} is the Laplacean of the image.

These geometrical invariants are considered over the scale range of an image, i.e. at a large number of different values of the Gaussian width σ. It is convenient to reparametrize σ as follows:

$$\tau = \ln \frac{\sigma}{\sigma_0}$$

where σ_0 is a scale of reference, which usually equals the pixel width of the image (i.e the inner scale). If features are considered at multiple scales, an equidistant sampling of τ is most natural.

To illustrate the importance of a multiscale analysis, consider figure 2, where pixels from the various segments are indistinguishable on the basis of local image properties, since the intensity ranges overlap. Upon using scaled features, thus taking the environment of a pixel into account, one might be able to discriminate between the pixels. It is important to remark that, although scaled operators are to be used, it is unknown beforehand at which scale a certain structure will appear. Another argument for a multiscale approach is that spatial connections of pixels are considered implicitly since feature patterns are likely to resemble for adjacent pixels, which favors the ability of an equal classification by the network. The issue of the network and the applied classification strategy is covered in the next section.

3 Classification algorithm

The purpose of our study is to segment digital images, in particular $2D$ and $3D$ medical images. The previous section shows how each pixel in an image can be assigned a feature pattern in terms of the relevant geometric invariants evaluated at that location. We assume that feature patterns of pixels which belong to one segment project to coherent and distinguishable regions in feature spaces. Furthermore we assume that the regions in feature spaces are robust in the sense that segments in similar images correspond to similar regions in the feature spaces. This allows us to take these regions as a basis for pixel classification.

The method we apply to determine the class regions is based on a type of self-organizing neural networks, the *Kohonen feature map* [Koh84]. Our approach requires one or a number of training image(s) (one may be sufficient because each image contains a large number of pixels), as well as a priori segmented versions of the training images.

Initially the values of a number of selected features are to be calculated for each pixel in the training image, to form the feature patterns. The features need to be normalized, since distance measures have to be evaluated in the feature space. We have chosen the normalization:

$$F' = \frac{F - \bar{F}}{\sigma_F}$$

where $\bar{F}$ is the average value for feature F in the image, and σ_F is the standard deviation of the feature in the image.

Feature patterns are repeatedly input of a Kohonen network, of which the architecture is illustrated in figure 1. The neurons in the so-called *map* have a predefined topological order, and the weight vector of each neuron is initialized randomly around the origin. Upon each input, the winning neuron c in the map is determined, whose weight vector has the smallest Euclidian distance to the current input pattern. The weight vectors of all neurons which are in a neighborhood N_c of the winner, according to the topological order of the map, are pulled into the direction of the input pattern. This gives the Kohonen learning rule:

$$\vec{w}_i(t+1) = \begin{cases} (1-\alpha)\vec{w}_i(t) + \alpha\vec{x}(t) & \text{for } i \in N_c \\ \vec{w}_i(t) & \text{otherwise} \end{cases}$$

where $\vec{w}_i(t)$ is the weight vector of neuron i at moment t, and α is the learning rate. Usually the learning rate decreases in time, as does the size of the neighborhood. So initially the global structure of the network evolves, while during further processing the values of the distinctive weight vectors are tuned finer. We opted for linear decrease of the extent of the neighborhood. We applied a local learning rate in each neuron, which is only lowered if the corresponding weight vector is updated. This causes a faster and more optimal organization of the weight vectors, than if a global learning rate is applied which is equal for all neurons [Ber88]. The learning rate decreased as follows:

$$\alpha_i(t+1) = \begin{cases} \frac{\alpha_i(t)}{1+\alpha_i(t)} & \text{if } \alpha_i(t+1) \geq \alpha_\infty \\ \alpha_\infty & \text{otherwise} \end{cases}$$

where α_i is the learning rate of neuron i. Typical values of the parameters are: $\alpha_i(0) = 0.5$ and $\alpha_\infty = 0.01$.

After a sufficient number of iterations the weight vectors of the network form a model of the feature space in terms of so-called *prototypical feature patterns*.

A slight modification of the algorithm is required since it may happen that there are weight vectors which do not match any input pattern, owing to the fact that weight vectors which are updated once are likely to match input patterns again. To avoid this problem a *conscience mechanism* may be added to the network [HN90, HKP91]. The basic concept of this mechanism is to keep track of the fraction of time that a neuron is the winner. A neuron will get a handicap if this fraction of time is substantially larger than $\frac{1}{N}$, where N is the total number of neurons. If the fraction of time is smaller than $\frac{1}{N}$, that neuron will be given preference to win. Define $f_i(t)$ as the fraction of time a neuron won until moment t. The next expression shows how the value of $f_i(t)$ can be calculated locally in each neuron:

$$\begin{aligned} f_i(0) &= 0 \\ f_i(t+1) &= (1-\beta)f_i(t) + \beta o_i(t) \end{aligned}$$

where $o_i(t)$ is the output of the neuron at moment t, which equals 1 if the neuron is the winner, and 0 otherwise. The constant β is usually chosen very small, a typical value for

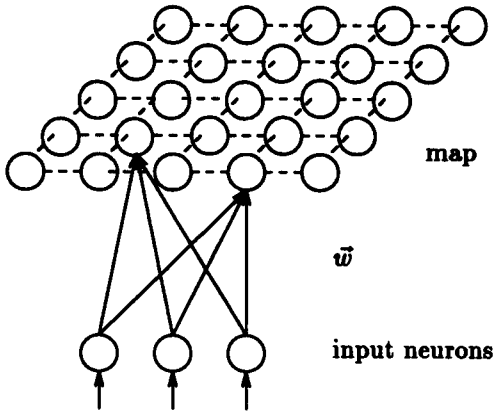

Figure 1: The architecture of a Kohonen network. Each input neuron is fully connected with the neurons in the map, but only a few connections are illustrated. The weight vectors $\vec{w}$ of the neurons in the map form the pivot of the network.

β is 10^{-4}. The handicap for neuron i can be calculated as follows:

$$h_i(t) = \gamma \left(\frac{1}{N} - f_i(t) \right)$$

where γ is a positive constant with a typical value of 10. The winner is the neuron which has the lowest value for the following expression:

$$d(\vec{w}_i, \vec{x}) - h_i$$

where d denotes the Euclidian distance.

Once we have obtained a satisfying set of prototypical feature patterns, the patterns may be labeled according the class they represent. It is straightforward to label each weight vector according the class of its best matching feature pattern, as derived from the a priori segmentation of the training image.

Each region in the feature space, corresponding to a single segment, is now represented by means of a number of equally labeled prototypical feature patterns. Classification of pixels is straightforwardly done by comparing the feature pattern representation of the pixel with all weight vectors, and assigning to the pixel the class of the best matching weight vector.

4 Results

Before we applied our method to medical images, we performed various tests on artificial images, which allow an objective evaluation of the segmentation results. The results obtained on medical images were evaluated by visual inspection since 'perfect' segmentations cannot be defined.

As concerns the network configuration, 100,000 iterations were executed with a 10×10 map, with inclusion of the conscience mechanism. Initially the value of the learning rate was 0.5, while its lower bound equaled 0.01. The extent of the neighborhood relation in the map was initially 7×7, decreasing 2 in each direction, every 1,000 iterations, until after 3,000 iterations weight vectors are updated individually.

For convenience we will denote $L_iL_i(4)$ for the feature L_iL_i obtained with $\sigma = 4$ etc. $L(0)$ will refer to the original intensity.

4.1 Intensity-image

The first tests we performed concerned images with mainly intensity information (instead of other, like texture, information). The image in figure 2*a* acted as the training image, with the image of figure 2*b* as the a priori segmentation. We applied the resulting network to segment the image in figure 2*c*. Although both images are much alike, the distribution of objects in both images is slightly different. Besides there is a severe intensity overlap among the various segments owing to noise. The standard deviation of the Gaussian noise was 200, while the difference between the mean intensities in the inner segment and 'ring' segment also amounted to 200. Results are given table 1 and table 2.

From the tables we may derive that the application of features at multiple scales supports the ability to classify the pixels. However, pixels at boundaries of segments remain difficult to classify on the basis of zeroth order information only. As expected the application of first order information improves the ability to classify pixels near boundaries, although the effect, which is clear from visual inspection, is not supported by the correctness measure. This discrepancy suggests that the measure is inappropriate for the purpose of evaluating this aspect of the segmentation. For this reason we did a number of additional evaluations, applying a criterion which discriminates between pixels near boundaries and pixels inside segments (table 3). These evaluations consent the visual inspections.

Second order information did not appear to be of importance in the considered images.

An example of a two-dimensional feature space is given in figure 3*a*, with features $L(0)$ and $L(4)$. Figure 3*b* reveals that the derived classes are not linearly separable.

$L(0)$	$L(\sigma)$ 1	2	4	8	%
•					72 ±4
•	•				86 ±2
•		•			84 ±2
•			•		88 ±2
•				•	86 ±2
•	•	•	•	•	96 ±1 *

Table 1: The numbers in the second row correspond to the scales of the features mentioned above. The 'bullets' explain which combination of features is taken into account, and results marked with a * are illustrated in a figure. In the column on the right the percentage of correct classified pixels is shown. The table shows that the application of zeroth order information at multiple scales is a favorable choice.

$L(0)$	$L(\sigma)$ 1	2	4	8	$L_iL_i(\sigma)$ 1	2	4	8	%
•					•				66 ±4
•						•			76 ±2
•							•		76 ±2
•								•	82 ±2
•	•	•	•	•				•	96 ±1
•	•	•	•	•	•	•	•	•	96 ±1 *

Table 2: This table comprises the results obtained using zeroth and first order information. A visual evaluation reveals that the application of first order information supports the ability to classify pixels near boundaries. The percentages in the tables do not support this since there are so few pixels at boundaries of segments. (See also table 3.)

$L(0)$	$L(\sigma)$ 1	2	4	8	$L_iL_i(\sigma)$ 1	2	4	8	% bound. pixels	% correct class. bound. pixels	% correct class. other pixels	total %
•	•	•	•	•					21.4	82 ±2	99.6 ±0.3	96 ±1
•	•	•	•	•	•	•	•	•	21.4	86 ±1	99.4 ±0.3	96 ±1

Table 3: In the above table different percentages are given for pixels near boundaries and pixels inside segments. We defined *boundary pixels* as pixels in whose direct environment of 3 × 3 pixels a change of object occurs. In the tenth column the percentage of boundary pixels in the image is given and the next column shows the percentage of correct classified boundary pixels.

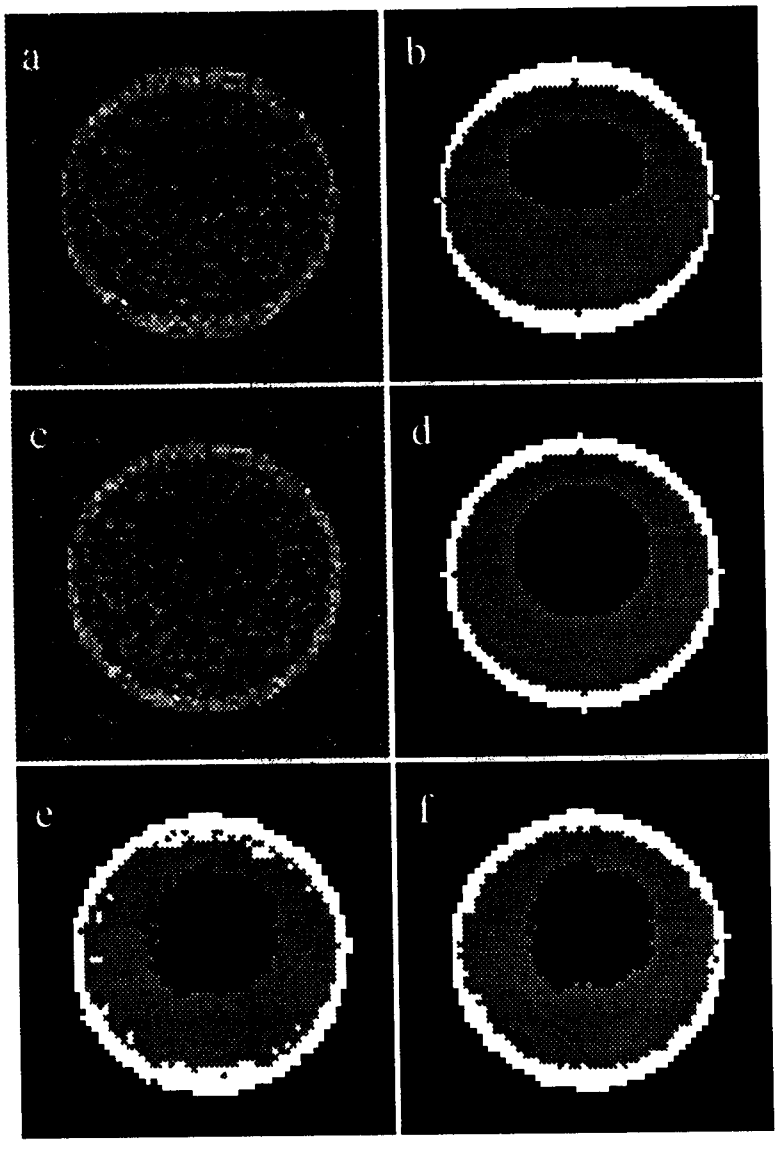

Figure 2: *a*) The training image.
b) The a priori segmentation of the training image.
c) A test image.
d) The required segmentation of the test image.
e) A segmentation of the test image on the basis of $L(0)$, $L(1)$, $L(2)$, $L(4)$, and $L(8)$.
f) A segmentation of the test image using the features $L(0)$, $L(\sigma)$ and $L_iL_i(\sigma)$, for $\sigma = 1, 2, 4, 8$.

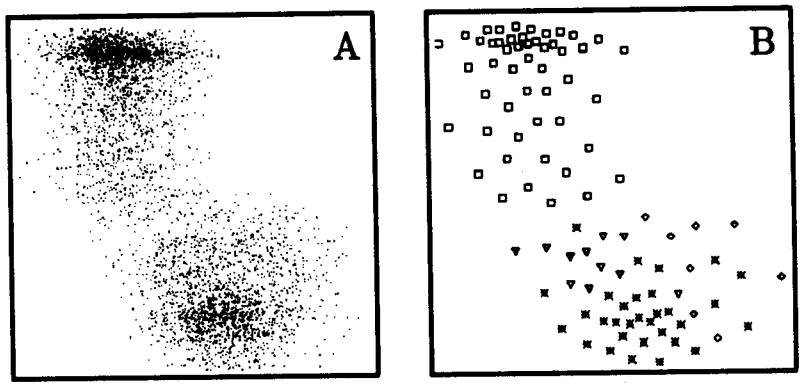

Figure 3: *a*) A feature space, acquired from the image in figure 2*a*. The features are $L(0)$ (vertically) and $L(4)$ (horizontally). Each dot corresponds to a feature pattern (pixel).

b) The values of the weight vectors of a map, which was trained with the patterns from *a*. Each weight vector is labeled in correspondence to the object it represents in order to delineate the class boundaries. Clearly the derived classes are not linearly seperable.

	$L(\sigma$			$L_iL_i(\sigma)$			$L_ii(\sigma)$			$L_{ij}L_{ji}(\sigma)$			$L_iL_{ij}L_j(\sigma)$			
$L(0)$	1	$\sqrt{2}$	2	1	$\sqrt{2}$	2	1	$\sqrt{2}$	2	1	$\sqrt{2}$	2	1	$\sqrt{2}$	2	%
•	•	•	•													80.0 ±0.5
•	•	•	•	•	•	•										93 ±2
							•	•	•							94 ±2
	•	•	•				•	•	•							94 ±2
				•	•	•	•	•	•							98 ±1
							•	•	•	•	•	•				98 ±1
							•	•	•				•	•	•	99 ±1
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	99.5 ±0.5 *

Table 4: This table comprises the percentages of correctly classified pixels in the segmentations of the texture-image.

4.2 Texture-image

We expected second order information to be important if the segments are composed of different textures. Two images were created, both consisting of three segments, only distinguishable on the basis of texture. Gaussian noise was added to both images. We attempted to segment the image in figure 4.2*b*, while the image in figure 4.2*a* acted as the training image. Since the images tend to 'flatten' if $\sigma > 2$, the maximum scale we considered equaled 2. Hence, the segmentations were based on fairly local image properties (as texture is). Results are given in table 4.

Obviously the use of zeroth order information only does not allow suitable segmentations of the image. It appears necessary to take higher order information into account. Second order information is of major importance in segmenting the texture-image.

4.3 MRI of the head

We have applied our strategy to segment various MR images of the head. Figures 5*a* and *b* illustrate the training image and a manually segmented version of it. A network was trained with feature patterns derived from the training image, and its weight vectors were labeled according the a priori segmentation. With this network several images, resembling to the training image, were segmented. The resulting segmentations are illustrated when zeroth order information at various scales was taken into account (figures 5*c-f*).

From the results we may conclude that the representations of the objects in the images are robust and fairly insensitive to shape and position variations, because a number of images with significant differing intensity distributions could be segmented satisfyingly with the same network.

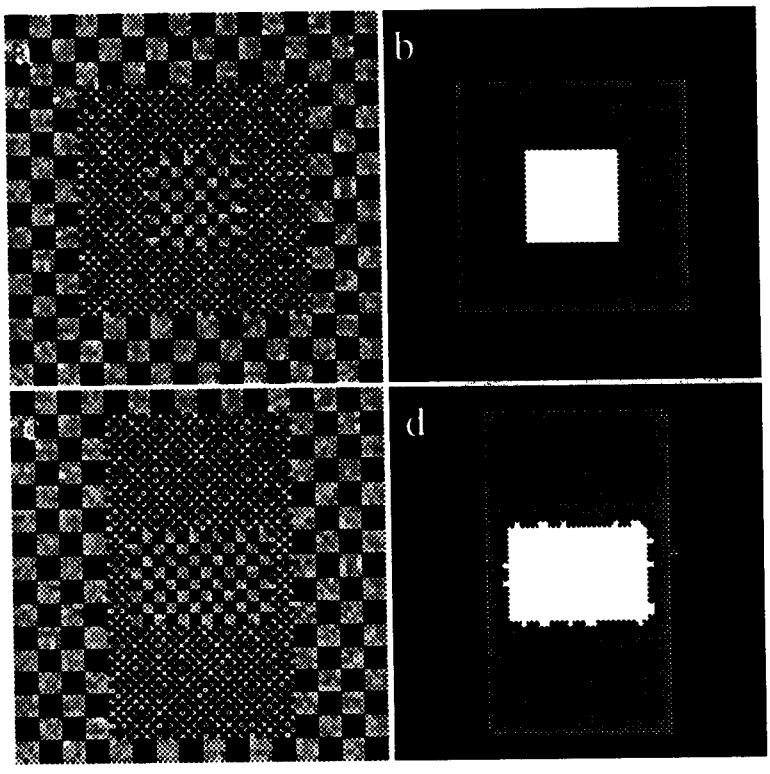

Figure 4: *a*) The training image.
b) The a priori segmentation of the training image.
c) A test image.
d) The automaticallly obtained segmentation of the test image, on the basis of $L(0)$, $L(\sigma)$, $L_iL_i(\sigma)$, $L_{ii}(\sigma)$, $L_{ij}L_{ji}(\sigma)$, and $L_iL_{ij}L_j(\sigma)$, $\sigma = 1, \sqrt{2}, 2$.

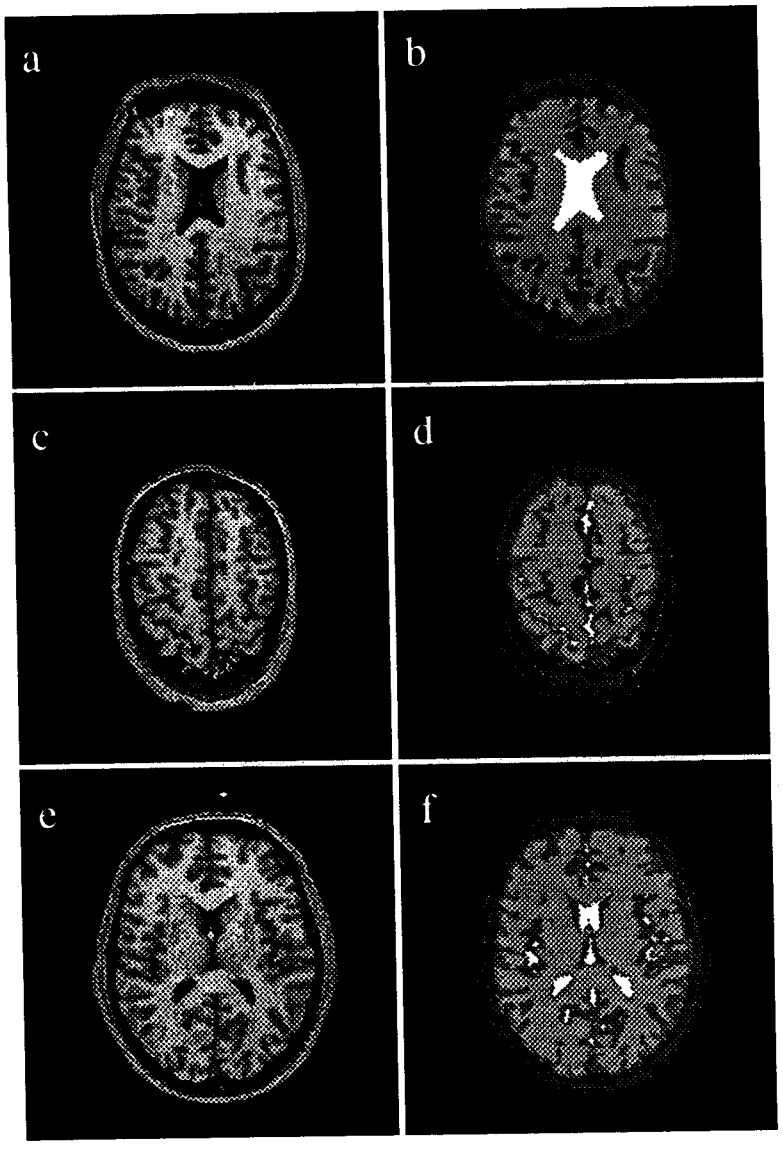

Figure 5: *a*) The training image.
b) The manually obtained segmentation of the training image.
c) A test image from the same three-dimensional dataset as the training image.
d) The acquired segmentation of the image in *c*. Using as features $L(0)$, $L(1)$, $L(2)$, $L(4)$, $L(8)$ and $L(16)$.
e) An image from a different dataset as the one the training image was obtained from.
f) The acquired segmentation of the image in *e*. Using as features $L(0)$, $L(1)$, $L(2)$, $L(4)$, $L(8)$ and $L(16)$.

5 Conclusion

A first conclusion of our study is that we have developed a system which is capable of segmenting similar images, once it has been trained to segment a representative one.

Apparently scaled differential invariant features provide a means of describing pixel properties, on which basis pixels can satisfyingly be classified. The multiscale approach ensures that the characteristics of pixels are not restricted to local properties, but extend to environments of varying extent. Applying features at multiple scales yields robust descriptions of the objects in the images, by feature patterns which are prototypical for the object, and supports the acquisition of homogeneous segments.

Inclusion of gradient information improves the classification of pixels near boundaries of segments, and second order information appears to be of importance for representing the characteristics of textured segments.

As far as we are concerned the Kohonen algorithm is a useful tool for image segmentation, although we did not compare the method with other classification strategies. A disadvantage of the Kohonen feature map is its probablistic character, so that the results are not exactly reproducible. We experienced difficulties with combining features, future investigations may reveal the underlying reasons. Another subject for further research is the application to three-dimensional images. We expect a better performance if the network is trained by feature spaces corresponding to three-dimensional images, since the characteristics of objects are essentially three-dimensional. We also propose to investigate the nature of the regions in feature spaces which correspond to segments, in order to be able to determine by means of which unsupervised strategies meaningful classes may be accomplished. An interesting option is to label each weight vector in such a way, that it indicates the probabilities a feature pattern (corresponding to a pixel) belongs to various segments. In such a manner probabilistic segmentations as pioneered in [VKV92] come within reach.

References

[BC91] L.C. Baxter and J.M. Coggins. Supervized pixel classification using a feature space derived from an artificial visual system. Technical report, University of North Carolina, Department of Computer Science, 1991.

[Ber88] H. Bertsch. Die selbstlernende topologische Merkmalskarte zur Bildsegmentierung und Klassifikation. Technical report, Deutsches Krebsforschungszentrum Heidelberg, Abteilung Medizinische und Biologische Informatik, 1988.

[FtHRKV92] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever. Scale and the differential structure of images. *Image and Vision Computing*, 10(6):376–388, July/August 1992. Special Issue: Information Processing in Medical Imaging.

[HKP91] J. Hertz, A. Krogh, and R.G. Palmer. *Introduction to the Theory of Neural Computation.* Addison-Wesley, 1991.

[HN90] R. Hecht Nielsen. *Neurocomputing.* Addison-Wesley, 1990.

[Koh84] T. Kohonen. *Self-Organization and Associative Memory.* Springer-Verlag, 1984.

[KvD84] J.J. Koenderink and A.J van Doorn. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.

[KvD90] J.J. Koenderink and A.J van Doorn. Receptive filed families. *Biological Cybernetics*, 63:291–297, 1990.

[tHRFKV91] B.M. ter Haar Romeny, L.M.J. Florack, J.J. Koenderink, and M.A. Viergever. *Scale-space: its Natural Operators and Differential Invariants*, volume 511 of *Lecture Notes in Computer Science*, pages 239–255. Springer-Verlag, 1991.

[VKV92] K.L. Vincken, A.S.E. Koster, and M.A. Viergever. Probabilistic multiscale image segmentation. In *Proceedings of the Second Conference on Visualization in Biomedical Computing.* IEEE, 1992.