Proc. NOBIM-konferansen 1994
Asker, Norway, June 1994
Note: Figure 3 in the original paper was wrong, the correct figure is substituted in this version

Segmentation of text/image documents using texture approaches

Trygve Randen and John Håkon Husøy
Rogaland University Center
Department of Electrical and Computer Engineering
P.O. Box 2557 Ullandhaug, N-4004 Stavanger, Norway
E-mail: tranden@hsr.no

Abstract – The digital computer and computer networks have made it possible to search for and retrieve electronically stored documents in seconds, no matter where in the world they are stored. This is far from the reality for documents stored as paper copies. Therefore there is considerable interest in digitizing paper documents. To digitize existing paper documents, it is of great importance to be able to separate the text from the graphics, in order to make the text searchable and more efficiently stored.

In this paper we present an approach to segmentation of text and graphics in scanned documents, based on the assumption that the text in a document may be viewed as one texture, while the graphics is a different texture. Using this assumption, we segment the documents with a texture segmentation scheme using filter banks as the feature extractors. While most traditional text-graphics segmentation schemes assumes some a priori knowledge of the input, our approach is independent of document layout, typeface, font size, scanning resolution etc.

Another approach to texture segmentation of documents for text-graphics segmentation has been presented by Jain and Bhattacharjee, using the Gabor filter as the feature extractor. In this paper we show that equally good results may be obtained using much more computationally efficient critically sampled perfect reconstruction filter banks.

1 Introduction

Traditionally, paper has been the main source for publication of newspapers, periodicals, reports, books etc. However, the availability and economy of powerful computers opens new possibilities for the storage, retrieval and search for electronically stored documents. An article located one place, may be retrieved from the other side of the world in seconds. In addition, if the text of the documents is stored in a suitable way, it is possible to search the contents of several documents in seconds.

However, it is not trivial to automatically digitize a document in a suitable manner. The straight-forward approach is to scan the whole document, and store it as a bit-map representation, but this approach requires a huge amount of storage capacity (even with modern image compression techniques) and does not enable search for text in the documents. The straight-forward approach will be applied for the proceedings of the IEEE 1994 International Conference on Acoustics, Speech & Signal Processing (ICASSP'94). These proceedings will be made available on CD-ROM, stored as TIFF bitmaps. Only the title and the abstract will be available in a searchable form.

A more efficient way of representing documents is to separate the text and the graphics, and store the text as ASCII text and the images as bit-maps. Therefore, several approaches to text-graphics segmentation has been proposed recently. The most popular approaches may broadly be classified as either *top-down* or *bottom-up* approaches [1]. The most common top-down techniques are the *run length smoothing* [2, 3, 4] (also known as the *constrained run length* method) and the *projection profiles* method [5, 6, 7, 8]. The major drawback with the top-down methods is that they are restricted to rectangular blocks, and therefore not suited to skewed text.

The bottom-up methods are typically variants of the *connected components* method [9, 10, 11]. The drawbacks with the connected components method is that it is character size dependent, sensitive to inter-line and inter-character spacings and sensitive to the digitizing resolution. It is, however, not restricted to rectangular blocks, as the top-down approaches.

A method different from the previous approaches, overcoming the limitations noted and not assuming any a priori knowledge of the document to be processed, was presented by Jain and Bhattacharjee [12]. The basic assumption underlying this approach is that the text regions of an image may be considered as an approximately uniform texture, and that images form another texture. The document may then be segmented by a texture segmentation scheme. The texture segmentation scheme used by Jain and Bhattacharjee [12] is mainly the same a the one proposed by Jain and Farrokhnia [13]; multichannel filtering with Gabor filters, the application of a local energy function and clustering by a non-supervised [14] classifier. The Gabor filter is the most frequently used filter for texture segmentation, and the rationale presented for the use of the Gabor filter is its relation to models for the early vision of mammals as well as its joint optimum resolution in time and frequency [15, 16].

One major drawback with the Gabor approach is its computational complexity. The approach presented by Jain and Bhattacharjee [12] uses a subset of eight filters, from the total set of 20 filters suggested by Jain and Farrokhnia [13], but even when this restricted subset is used, a total data amount of eight times the input has to be used as features for the classifier. If it is possible to reduce the feature set, the computational complexity of the segmentation will be reduced.

It is well known that sub-sampling of the various frequency bands, in such a way that the total number of samples in the sub-bands equals the number of signal samples in the input image, can be performed without losing any information, as long as the filters are designed appropriately [17]. We have previously shown that such a set of filters may be successfully used for a general segmentation approach [18], with approximately the same segmentation quality as with the Gabor filters. With this set of filters, *considerable* computational savings were realized. In this paper we address the segmentation of document images with this set of filters, using a non-supervised classifier. In Section 2 we describe the general texture segmentation setup, mainly as suggested in [18], and in Section 3 we present the results of our experiments on different images under varying conditions.

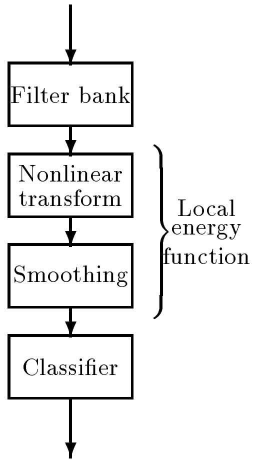


Figure 1: Block diagram for the basic experimental setup.

2 Multichannel filtering for texture segmentation

Experiments done on cats and apes, which have a visual system similar to the human visual system [19], show that these animals decompose the visual images into several filtered images, with different spatial frequency ranges and tuned to different orientations [20]. This model is used for our texture segmentation scheme, and Figure 1 shows our general experimental setup. In the following subsections, we will describe the elements of this setup.

2.1 Filter bank

The basic idea underlying the multichannel filtering for texture segmentation is that texture is described by its spatial-frequency content. By choosing a *perfect reconstruction* filter bank, it is possible to subsample the output from the filter bank such that the total number of signal samples on the output is equal to the number of samples on the input, without any loss of information. Remember that the output from *each* Gabor filter in a filter bank of Gabor filters had a number of signal samples equal to the number of samples on the input. Critically sampled filter banks with perfect reconstruction, have been extensively used in image cod-

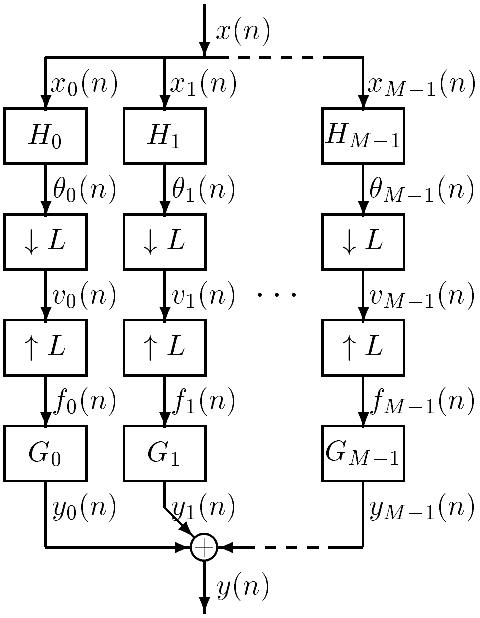


Figure 2: General analysis-synthesis filter bank.

ing applications, see for example Husøy [21], and we have previously shown their applicability to general texture classification [22] and segmentation [18].

We start with a short description of critically sampled filter banks, as introduced in the image coding literature. Figure 2 shows a general *analysis - synthesis* filter bank. The H_i filters are the analysis filter bank, and the G_i filters are the synthesis filter bank. The symbol $\downarrow L$ represents *sub-sampling* by L, while the symbol $\uparrow L$ represents *up-sampling* by L. Sub-sampling by L is the process of discarding all but every L'th sample, while up-sampling by L is the process of inserting $L-1$ zeros between each sample.

The filter bank in Figure 2 is *critically sampled* if $L = M$, that is the sub-sampling factor is equal to the number of filters in the filter bank. A critically sampled filter bank has *perfect reconstruction* if the output of the synthesis filter bank is equal to the input to the analysis filter bank (with a possible delay). There exists a large (in fact infinite) class of filter banks with perfect reconstruction.

It is possible to implement the filtering operations of the filter banks at low sampling rate, and thereby achieve big savings in computational complexity [23]. The theory of filter

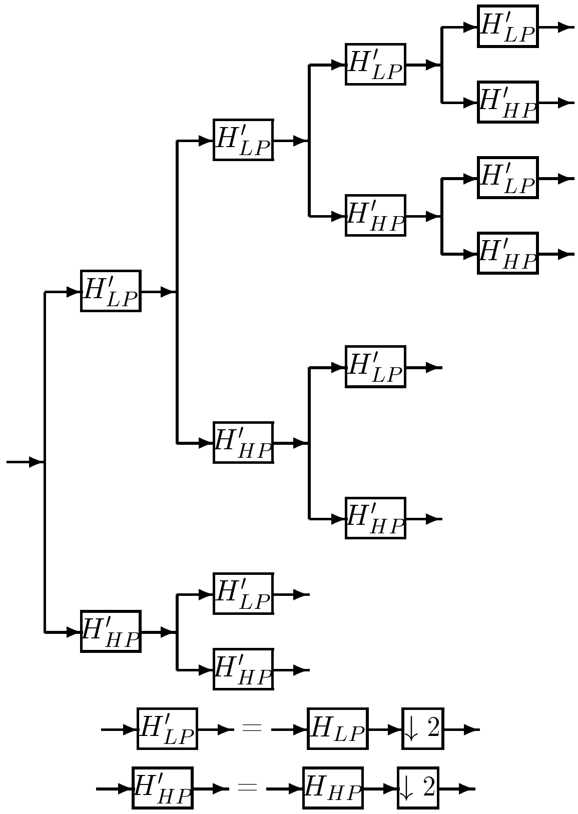


Figure 3: Tree-structured filter bank.

banks is covered in detail in [17, 23].

A filter bank is *separable* if it is possible to apply the filters in each dimension separately. A filter bank is *tree-structured* if it is built up as a hierarchical tree of filters, applied recursively to the outputs of the filters in the previous level. We restrict our further discussion to the separable filter banks with a structure in each dimension as shown in Figure 3. The filters $H_{LP}(z)$ and $H_{HP}(z)$ are low-pass and high-pass filters, respectively, with pass band width of $\pi/2$.

We will use the the infinite impulse response (IIR) filter bank 'F_2_1_09' from [21]. The choice of 'F_2_1_09' is based on a large number of experiments on texture segmentation with various perfect and nearly perfect reconstruction filter banks [24]. The IIR filters are based on first order digital *all-pass* filters, and they are given by

$$
\begin{aligned}
H_{LP}(z) &= A_0(z^2) + z^{-1}A_1(z^2) \qquad (1)\\
H_{HP}(z) &= A_0(z^2) - z^{-1}A_1(z^2), \qquad (2)
\end{aligned}
$$

where

$$A_i(z) = \frac{a_i + z^{-1}}{1 + a_i z^{-1}},\ i = 0, 1$$

are the all-pass filters. For 'F_2_1_09' the coefficient $a_0 = 0.66$, while a_1 may be zero or one, depending on the implementation.

We have already commented on the computational attractiveness of the critically sampled filter banks. In addition, the IIR filters, as opposed to finite impulse response (FIR) filters, are especially attractive due to their computational simplicity [21]. Thus, our approach to texture segmentation with multichannel filtering is very attractive due to the low computational complexity.

2.2 Local energy function

The objective of the multichannel filtering, and the subsequent local energy function, is to transform the edges between the textures into detectable discontinuities [25]. The filter bank is a set of bandpass filters with frequency and orientation selective properties and the local energy function is utilized for the purpose of transmitting areas in each channel where the band pass frequency components are strong into a high constant gray-level and areas where they are weak into a low constant gray-level.

We found the local energy function proposed by Jain and Farrokhnia [13] to be appropriate for our experiments, with minor modification for multi rate processing. Thus we use

$$\left|\tanh\left(\alpha r_k(x, y)\right)\right|, \tag{3}$$

as the nonlinear transform, where $r_k(x, y)$ is the image corresponding to frequency channel no. k and (x, y) denote spatial coordinates. The parameter α gives the saturation characteristic of this function. By experience, $\alpha = 0.25$ [13].

The nonlinear transform is followed by a Gaussian low-pass (smoothing) filter of the form

$$H_G(u, v) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(u^2+v^2)}. \tag{4}$$

The parameter σ determines the pass-band width of this low-pass filter. For a non-subsampled filter-output we use

$$\sigma = \frac{\sqrt{2}}{2}\frac{1}{f_0},$$

where f_0 is the radial center frequency from the band-pass filter, normalized to the interval $[-1/2, 1/2]$. When we decimate the signal by a factor N, a sinusoid with period M and frequency $f = 1/M$ will get the period M/N, and the frequency $N/M = Nf$. Thus, the smoothing filter designed for a given full rate signal will no further be designed for the same signal after sub-sampling. We remember that σ is is inversely proportional to the radial frequency f_0, therefore we use the equation

$$\sigma = \frac{\sqrt{2}}{2}\frac{1}{Nf_0},$$

where N is the sub-sampling factor and f_0 is the absolute value of the center frequency of the filter bank channel.

2.3 Classification

The output from the local energy function is a set of images, one image per filter in the filter bank. These images are the feature images, that will form the basis for the classification. We use these images to form feature vectors, where each feature image corresponds to one element in the feature vectors. Classification is the task of assigning these feature vectors to different classes.

A supervised classifier for document image segmentation would mean that the segmentation would be dependent on scale, rotation and font. One goal with our approach is to make a segmentation scheme independent of these issues, thus we need a non-supervised classifier. In our experiments we use the well-known *c-means* clustering algorithm [14].

3 Experimental results

We have chosen to use parts of pages scanned from the march 2., 1994 issue of "Stavanger Aftenblad" (SA) as our test images. For comparison purposes, we have also included some images used in [12, 26]. All the images used in this paper are available via anonymous FTP from `ftp.hsr.no`.

Figure 4 shows a part of page 8 in SA, and the successful two-class segmentation of the image. Figure 5 shows the segmentation results from the same page, rotated by 55° and at half


gir en egen stemning

billig og miljøvennlig – ved er dyrer

I oktober i fjor kjøpte Kjell og Olaug Herigstad på Storhaug to favner med ved. De regner med å bruke opp nesten alt før vinteren og kulden slipper tak.

a

a

b

b

Figure 4: The 1024 × 1024 pixel scan of parts of page 8 from SA (a), and the segmentation of the page (b).

Figure 5: Segmentation of page 8 from SA rotated by 55° (a), and segmentation of the page scanned at half the resolution (b), 512 × 512 pixels.

a

b

Figure 6: Test image used by Jain and Bhattacharjee (a), and segmentation of the image (b).

the resolution, respectively. This figure shows one important result: the segmentation quality is invariant to scaling and rotation, except that the results from the headings are somewhat more inconsistent and parts of the graphics of the rotated image are classified as text.

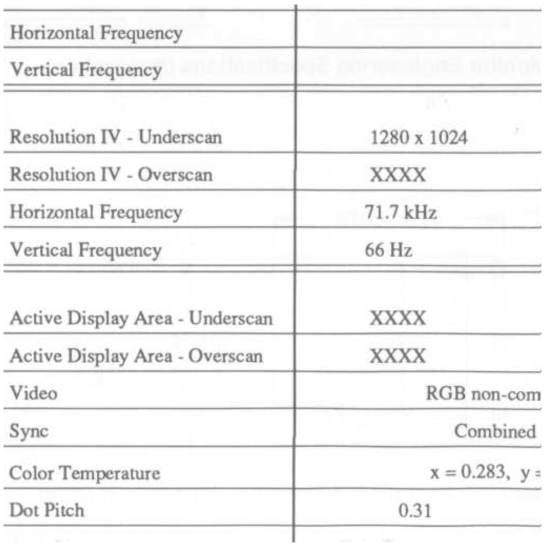
Figures 6 and 7 are the same images as used in [12] and [26]. The segmentation results from these images are practically the same, verifying the conclusion in [18] that our new filtering approach, including subsampling, is appropriate.

Our text segmentation method is based on the assumption that the text regions may be considered as constituting a different texture than the picture regions. Therefore it is natural that the method performs poorer when the images contain textured regions similar to the texture in the text region, like in Figure 8(a). We see that the two-class segmentation, Figure 8(b), in this case has included a considerable part of the graphics in the text category. We did experiments with three classes, and as Figure 8(c) shows, this leads to a somewhat better segmentation, where most of the figure areas that were misclassified as text in the two-class experiment now fall in the third class.

It is important to be aware that all our experiments were done with no assumptions on the input image and no post-processing. Several kinds of post-processing techniques and modifications to the scheme are possible, and should improve the results, especially if we know more about the format of the input. If we know the font used, it is possible to use a supervised classifier, tuned to detect that particular font. If the input is known to be restricted to standard newspaper columns, a post-processing technique looking for blocks of the particular width may be employed, and so on.

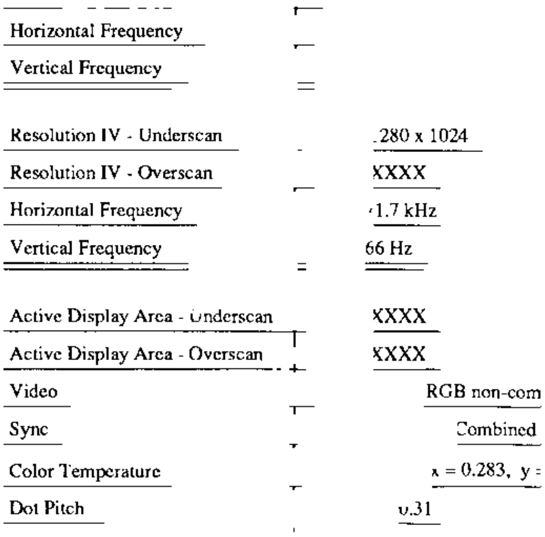
4 Summary and conclusions

Storage of documents in digital form is a useful tool in making the contents easily available, and if important documents are made available for search and retrieval by computers, it will be of great importance in the work of scientists, engineers, journalists etc. In digitizing existing paper documents, it is of great importance to be able to separate the text from the graph-



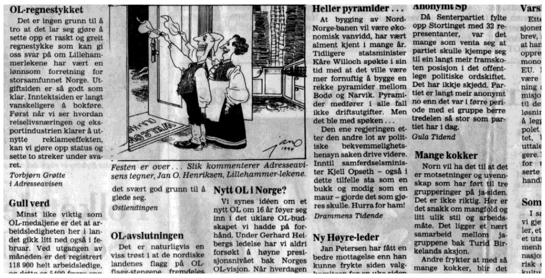
Horizontal Frequency	
Vertical Frequency	
Resolution IV - Underscan	1280 x 1024
Resolution IV - Overscan	XXXX
Horizontal Frequency	71.7 kHz
Vertical Frequency	66 Hz
Active Display Area - Underscan	XXXX
Active Display Area - Overscan	XXXX
Video	RGB non-com
Sync	Combined
Color Temperature	x = 0.283, y =
Dot Pitch	0.31

a

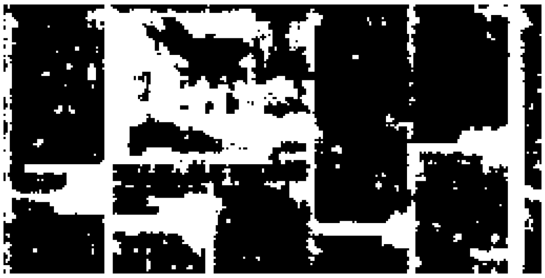


b

Figure 7: Test image used by Jain and Bhattacharjee (a), and segmentation of the image (b).



a



b

c

Figure 8: Page 2 from SA (a), two-class segmentation of the page (b), and three-class segmentation (c), areas with gray background are the third class.

ics, both because it is possible to store the text more effectively than the graphics, and because it is practically impossible to search for text in a bit-map file.

In this paper we have presented a new approach to segmentation of text and graphics in digitized document images, based on texture segmentation with filter banks. While most traditional text-graphics segmentation schemes assume some a priori knowledge of the input, our approach performed well even without any a priori knowledge of the input format.

References

[1] S. N. Srihari, "Document image understanding," in *Proc. IEEE Computer Society Fall Joint Computer Conf.*, (Dallas, TX), pp. 87–96, Nov. 1986.

[2] F. M. Wahl, K. Y. Wong, and R. G. Casey, "Block segmentation and text extraction in mixed text/image documents," *Computer Graphics and Image Processing*, vol. 20, pp. 375–390, 1982.

[3] F. Shih, S.-S. Chen, D. Hung, and P. Ng, "A document segmentation, classification and recognition system," in *Proc. Int. Conf. on Systems Integration*, pp. 258–67, 1992.

[4] P. Chauvet, J. Lopez-Krahe, E. Taflin, and H. Maître, "System for an intelligent office document analysis, recognition and description," *Signal Processing*, vol. 32, no. 1-2, pp. 161–190, 1993.

[5] O. Iwaki, H. Kida, and H. Arakawa, "A segmentation method based on office document hierarchical structure," in *Proc. Int. Conf. Syst. Man Cybern.*, (Alexandria, VA), pp. 759–763, 1987.

[6] D. Wang and S. N. Srihari, "Classification of newspaper image blocks using texture analysis," *Computer Vision, Graphics and Image Processing*, no. 47, pp. 327–352, 1989.

[7] G. Nagy, S. Seth, and M. Viswanathan, "A prototype document image analysis system for technical journals," *Computer*, vol. 25, no. 7, pp. 10–22, 1992.

[8] M. Krishnamoorthy, G. Nagy, S. Seth, and M. Viswanathan, "Syntactic segmentation and labeling of digitized pages from technical journals," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 7, pp. 737–747, 1993.

[9] L. A. Fletcher and R. Kasturi, "A robust algorithm for text string separation from mixed text/graphics images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 10, Nov. 1988.

[10] T. Saitoh and T. Pavlidis, "Page segmentation without rectangle assumption," in *Proc. Int. Conf. on Patt. Rec.*, pp. 277–80, Int. Assoc. for Patt. Rec., 1992.

[11] F. Lebourgeois, Z. Bublinski, and H. Emptoz, "A fast and efficient method for extracting text paragraphs and graphics from unconstrained documents," in *Proc. Int. Conf. on Patt. Rec.*, pp. 272–6, Int. Assoc. for Patt. Rec., 1992.

[12] A. K. Jain and S. Bhattacharjee, "Text segmentation using Gabor filters for automatic document processing," *Machine Vision and Applications*, vol. 5, no. 3, pp. 169–184, 1992.

[13] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recognition*, vol. 24, no. 12, pp. 1167–1186, 1991.

[14] R. Schalkoff, *Pattern Recognition: Statistical, Structural and Neural Approaches*. New York: Wiley, 1992.

[15] D. Gabor, "Theory of communication," *J. Inst Elect. Engr.*, vol. 93, pp. 429–457, 1946.

[16] J. G. Daugman, "Uncertainty relation for resolution in space, spatial-frequency, and orientation optimized by two-dimensional visual cortical filters," *Journal of Optical Society of America*, vol. 2, no. 7, pp. 1160–1169, 1985.

[17] P. P. Vaidyanathan, "Quadrature mirror filter banks, M-band extensions and perfect-reconstruction techniques," *IEEE ASSP Magazine*, vol. 4, pp. 4–20, July 1987.

[18] T. Randen and J. H. Husøy, "Multichannel filtering for image texture segmentation," *Optical Engineering*, vol. 33, pp. 2617–2625, Aug. 1994.

[19] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674–693, July 1989.

[20] F. W. Campbell and J. G. Robson, "Application of fourier analysis to the visibility of gratings," *J. Physiology*, vol. 197, pp. 551–566, 1968.

[21] J. H. Husøy, "Low complexity subband coding of still images and video," *Optical Engineering*, vol. 30, pp. 904–911, July 1991.

[22] J. H. Husøy, T. Randen, and T. O. Gulsrud, "Image texture classification with digital filter banks and transforms," in *Proc. SPIE International symposium on optical instrumentation and applied science: Applications of Digital Image Processing XVI*, (San Diego), pp. 260 – 271, July 1993.

[23] A. N. Akansu and R. A. Haddad, *Multiresolution Signal Decomposition*. San Diego: Academic Press, 1992.

[24] T. Randen, "Flerkanal filtrering for tekstursegmentering," Master's thesis, Rogaland University Center, 1993 (in Norwegian).

[25] D. F. Dunn, W. E. Higgins, and J. Wakeley, "Determining Gabor-filter parameters for texture segmentation," in *Intelligent robots and computer vision XI*, vol. 1826, pp. 51–63, 1992.

[26] M. Tuceryan and A. K. Jain, "Texture analysis," in *Handbook of pattern recognition and computer vision* (C. H. Chen, L. F. Pau, and P. S. P. Wang, eds.), ch. 2, pp. 235–276, Singapore: World Scientific, 1993.