Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation

Martin BAATZ und Arno SCHÄPE

Dieser Beitrag wurde nach Begutachtung durch das Programmkomitee als „reviewed paper“ angenommen.

Abstract

A necessary prerequisite for object oriented image processing is successful image segmentation. The approach presented in this paper aims for an universal high-quality solution applicable and adaptable to many problems and data types. As each image analysis problem deals with structures of a certain spatial scale, the average image objects size must be free adaptable to the scale of interest.
This is achieved by a general segmentation algorithm based on homogeneity definitions in combination with local and global optimization techniques. A scale parameter is used to control the average image object size. Different homogeneity criteria for image objects based on spectral and/or spatial information are developed and compared.

1 Motivation

Image analysis implies to deal with image semantics. In most cases important semantic information to understand an image is not represented in single pixels but in meaningful image objects and their mutual relations. Furthermore many types of image data are more or less textured. Airborne data, radar or VHR-satellite data are playing an increasing role in remote sensing. In most cases, analysis of such textured data can only be successful when they are segmented in meaningful ‘homogenous’ areas. As a coarse rule of thumb the scale of such image objects must be significantly larger than the scale of image noise respectively texture.
One application of object oriented image analysis is multi source data fusion. Integration of different data types plays an important role in the field of remote sensing and GIS integration. Given a set of georeferenced data of different and arbitrary origin the definite topology of image objects allows to bring these different types of data in a concrete local relation. For instance image objects can be extracted based on one data type. In subsequent analysis steps the image objects are able to take into account the attributes in other data layers. Also the classified image objects are a useful link on which remote sensing and GIS integration can be build on.

2 A Short Review on Image Segmentation

Procedures for image segmentation are a main research focus in the area of image analysis since years. Many different approaches have been followed. However, few of them lead to qualitatively convincing results which are robust and under operational settings applicable. Often, the expectation to a segmentation result is to automatically extract all objects of interest in an image concerning a certain task. This strategy oversees the considerable semantic multitude which in most cases needs to be handled to come to such a final result. Or it leads to the development of highly specified algorithms applicable only to a reduced class of problems and image data.

One of the simplest approaches to segmentation are all types of global thresholding. Typically they are leading to results of a relatively limited quality. Region growing algorithms are clustering pixels starting on a limited number of single seed points. These algorithms basically depend on the set of given seed points, often suffering from a lack of control in the break off criterion for the growth of a region. In many operational applications different types of texture segmentation algorithms are used. They typically obey a two-stage scheme [GEMAN & al. 1990, JAIN & FARROKHNIA 1991, MAO & JAIN 1992, HOFMANN & al. 1998] :

1. *In the modeling stage characteristic features are extracted from the textured input image which range from spatial frequencies* [BOVIK & al. 1990, JAIN & FARROKHNIA 1991, HOFMANN & al. 1998], *MRF-models* [DERIN & COLE 1986, MANJUNATH & CHELLAPPA 1991, MAO & JAIN 1992, WON & DERIN 1992, PANFWANI & HEALEY 1995], *co-ocuurence matrices* [HARALICK & al. 1973] *to wavelet coefficients* [SALARI & ZING 1995], *wave packets* [LAINE & FAN 1996] *and fractal indices* [CHAUDHURI & SARKAR 1995].
2. *In the optimization stage features are grouped into homogeneous segments by minimizing an appropriate quality measure. This is most often achieved by a few types of clustering cost functions* [GEMAN & al. 1990, JAIN & FARROKHNIA 1991, MAO & JAIN 1992, HOFMANN & al. 1998, MANJUNATH & CHELLAPPA 1991, WON & DERIN 1992, PANFWANI & HEALEY 1995, SHI & MALIK 1997] (citation: PUZICHA, BUHMANN, 1998). A further possibility is the watershed transformation [Wegner & al 1997].

Although texture segmentation leads to reproducible, for specific applications excellent and sometimes high speed results, they are mostly just applicable for a limited number of types of image data, texture types and problems. A further alternative are knowledge based approaches. They are trying to incorporate knowledge derived from training areas or other sources into the segmentation process [Gorte 1998]. These approaches deliver classified regions. Mostly they are specific, not necessarily robust and do not necessarily deliver homogeneous areas.

Alternatively a segmentation method was developed which beneath the spectral and textural properties of the objects to be detected also takes into account their different size respectively. their different behavior on different stages of scale.

2 Design Goals

The method presented in this paper is used to create object primitives as the first processing step in the object orientated image analysis software eCognition. The resulting image objects are the raw material for further classification and refinement procedures.

High Quality Image Object Primitives: The primitives should be an universal high-quality solution applicable and adaptable to many problems and even textured image data of arbitrary type.

Multiresolution: Objects of interest typically appear on different scales in an image simultaneously. The extraction of meaningful image objects needs to take into account the scale of the problem to be solved. Therefore the scale of resulting image objects should be free adaptable to fit to the scale of task.

Similar Resolution: Almost all attributes of image objects – color, texture or form - are more or less scale-dependent. Only structures of similar size respectively scale are of comparable quality. Hence the size of all resulting image objects should be of comparable scale.

Reproducibility: Segmentation results should be reproducible

Universality: Applicable to arbitrary types of data and problems

Speed: Reasonable performance even on large image data sets

3 Criteria for the evaluation of segmentation results

3.1 Quantitative Criteria

Given

- a certain definition of heterogeneity for image objects and
- a certain average size of image objects (e.g. in pixels)

possible criteria are:

(1) average heterogeneity of image objects should be minimized; or

(2) average heterogeneity of image objects weighted by their size in pixel should be minimized.

In this contribution preference is given to the second criterion. Without any change it can be formulated as:

(3) average heterogeneity of pixels should be minimized. Each pixel is weighted with the heterogeneity of the image object to which it belongs.

This formulation emphasizes the requirement that each arbitrary area of an image needs to be taken into account in a similar way compared to any other. There must be a symmetric contribution of each area to the average heterogeneity of the segmentation result.

3.2 Qualitative Criteria

A strong and experienced source for evaluation of segmentation techniques is the human eye. As segmentation procedures are used for automation of image analysis applications they are replacing the activity of visual digitizing. No segmentation result - even if quantitatively proofed - will convince, if it does not satisfy the human eye. One important prerequisite to reach this goal is the consistent handling of local contrasts. Another is the segmentation of image regions of a more or less similar dimension. A further important criterion for evaluation is the information which can be extracted from image objects for further successful processing.

4 General Concept

The procedure for the multi-scale image segmentation presented here can be described as a region merging technique. It starts with each pixel forming one image object or region. At each step a pair of image objects is merged into one larger object. The merging decision is based on local homogeneity criteria, describing the similarity of adjacent image objects. The homogeneity criterion is not only a 'fit' or 'fit not' criterion. A "merging cost" is assigned to each possible merge. These costs represent the *degree of fitting*. For a possible merge the *degree of fitting* is evaluated and the merge is fulfilled, if it is smaller than a given 'least degree of fitting'. The procedure stops when there are no more possible merges. A small 'least degree of fitting' permits fewer merges than a larger one. Therefore the size of the resulting image objects will grow with the 'least degree of fitting' value. Due to this property, the parameter will be termed in the following as the *scale parameter*. A merge with a better degree of fitting (i.e. smaller value) than the scale parameter is said to "*fulfill the homogeneity criterion*". The two main components of multiresolution segmentation are:

- Decision Heuristics to determine the image objects that will merge at each step
- Definition of a homogeneity of image objects to compute the *degree of fitting* for a pair of image objects

5 Decision Heuristics

Given a definition for image object homogeneity (respectively heterogeneity) a method for finding the image objects for a merge is required. Starting with an arbitrary object *A*, different heuristics can be applied to find an adjacent object *B* for the merge. The following four possibilities are distinguished by an increasing constraint in terms of freedom of choice:

Fitting: Merge *A* with any arbitrary neighbor object *B* for which the homogeneity criterion is fulfilled.

Best Fitting: Merge *A* with that neighbor object *B* with which the homogeneity criterion is fulfilled best. 'Best' means the homogeneity is fulfilled concerning the *scale parameter* and the merge has the smallest degree of fitting value compared to all possible merges of A with any other adjacent object *B'*.

Local mutual best fitting: For *A* find its neighbor object *B* with which the homogeneity criterion is fulfilled best. Find for *B* its neighbor object *C* with which *B* fulfills the homogeneity criterion best. Confirm that the homogeneity criterion is best fulfilled mutually ($C = A$). If not, repeat the same loop taking *B* for A and C for B. This heuristic allows to find the best fitting pair of objects in the local vicinity of *A* following the gradient of homogeneity.

Global mutual best fitting: Merge that pair of adjacent objects in the whole scene which fulfills the homogeneity criterion best. This heuristic does not follow a distributed treatment order. Instead the order is implicitly given by the heuristics.

Distributed treatment order of image objects: Except for global mutual best fitting each decision heuristics needs a given image object *A* as starting point for the search of the merging pair. For the maintenance of a similar size/scale of all image objects it is necessary to let them grow in a simultaneous way. This can be achieved by choosing a sequence of starting points which fulfills the following two conditions (for each object taken into account the procedure performs one merge):

1. Handle each point respectively each object once per cycle
2. Distribute subsequent merges as far as possible from each other over the whole scene.

A simple approximate solution for this problem is a procedure which handles each segment in the scene in a random sequence. It turns out that such a random sequence is sub optimal due to the percolation patterns or clusters which increasingly appear, starting at a certain density of points. Better procedures are sequences which are systematically taking points with maximum distance to all other points treated before. Such a procedure can for instance be derived from a dither matrix produced by a binary counter (Fig.1).

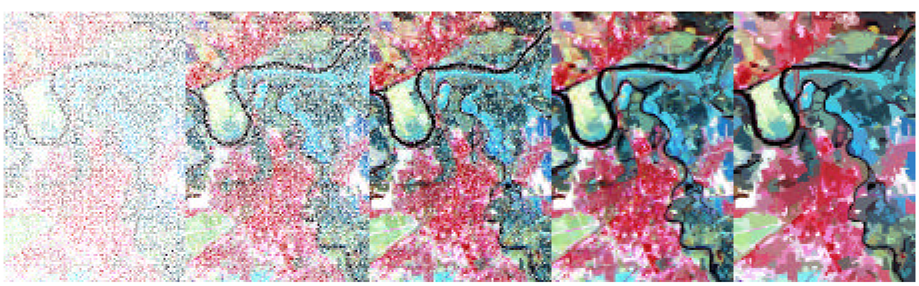

Fig.1: Different states of a multiresolution segmentation of LandSat data using a distributed treatment order in form of a binary counter over a dither matrix. The growth of image objects is regular over the whole image. Adjacent image objects are always of more or less similar scale.

The distributed treatment order can be defined over pixels or over segments. If the order is defined over segments each object will be treated once per cycle. If the order is defined over pixels then for each pixel the object is treated to which the pixel belongs. Hence larger objects are treated more often per cycle than small objects, taking into account the larger area they are representing.

6 Definitions for the ‘degree of fitting’

6.1 Difference between adjacent objects

Given a certain feature space, two image objects are similar which are near to each other in this feature space. For an *d*-dimensional feature space the *degree of fitting h* reads as

(4) $$h = \sqrt{\sum_d (f_{1d} - f_{2d})^2}$$

Examples for appropriate object features are for instance mean spectral values or texture features, such as the variance of spectral values. The distances can be furthermore standardized by the standard deviation over all segments of the feature in each dimension:

(5) $$h = \sqrt{\sum_d \left(\frac{f_{1d} - f_{2d}}{s_{fd}} \right)^2}$$

6.2 Change of heterogeneity in a virtual merge

Starting with single pixels objects and merging them pairwise to larger objects at least spectral average heterogeneity of all image objects obviously will increase. Basic goal of an optimization procedure must be to minimize the incorporated heterogeneity at each single merge. An image object should be merged to that adjacent image object that yields a minimum increase of heterogeneity.
Therefore we define the *degree of fitting* of two adjacent image objects by describing the change of heterogeneity h_{diff} in a virtual merge. Given an appropriate definition of heterogeneity for a single image object the increase of heterogeneity in a merge should be minimized. There are different possibilities to describe the change of heterogeneity h_{diff} before and after a virtual merge:

(6) $$h_{diff} = h_m - \frac{h_1 + h_2}{2}$$

This definition corresponds to the requirement of a quantitative criterion for the evaluation of a segmentation result given in (1). It can be improved by taking into account the objects size *n*:

(7) $$h_{diff} = h_m - \frac{h_1 n_1 + h_2 n_2}{n_1 + n_2}$$

An alternative is to weight image object heterogeneity with the object size, thus fulfilling the quantitative criterion in (2):

(8) $$h_{diff} = (n_1 + n_2)\, h_m - (n_1 h_1 + n_2 h_2) = n_1(h_m - h_1) + n_2(h_m - h_2)$$

This definition can be generalized to an arbitrary number of channels c, each having a weight w_c:

(9) $$h_{diff} = \sum_c w_c \left(n_1 \left(h_{mc} - h_{1c} \right) + n_2 \left(h_{mc} - h_{2c} \right) \right)$$

Appropriate definitions for spectral heterogeneity of image objects are for instance

(10) the variance of spectral mean values or

(11) the standard deviation of spectral mean values.

6.3 Form Heterogeneity

In the following two possible definitions for form heterogeneity are introduced. They can be incorporated into the spectral homogeneity. One is the deviation from the ideal compact form given by the relation between factual edge length *l* and the root of the object size *n* in pixels, i.e. the edge length of a square with n pixels.

(12) $$h = \frac{l}{\sqrt{n}}$$

Another possibility to describe form homogeneity is the deviation from the shortest possible edge length given by the bounding box *b* of the segment. It is the relation between factual edge length *l* and the edge length of the bounding box. In a raster the edge length of the bounding box is also the shortest possible edge length for an arbitrary segment, i.e. $l \geq b$ holds for any image object.

(13) $$h = \frac{l}{b}$$

7 Discussion

Global mutual best fitting is the strongest constraint for the optimization problem and it reduces heterogeneity most over the scene following a pure quantitative criterion. However, there is a significant disadvantage: It does not use the distributed treatment order and – in connection with a heterogeneity definition for color - builds first segments in regions with a low spectral variance. This leads to an uneven growth of image objects over a scene and to an unbalance between regions of high and regions of low spectral variance. The quantitative difference of the result to *local mutual best fitting* is neglectable. *Local mutual best fitting* always performs the most homogeneous merge in the local vicinity following the gradient of the *degree of fitting*. This happens even if the region is of relative high spectral variance. Together with a distributed treatment order, the growth of image objects happens simultaneously as well in regions of low spectral variance as in regions of high spectral variance. Starting with an arbitrary image object the *local mutual best fitting* heuristics switches in average 3 to 6 times to adjacent image objects until a pair of mutual best fitting objects was found. A distributed treatment order together with the *local mutual best fitting* guarantees a symmetric handling of regions of different texture and highly minimizes heterogeneity of resulting image objects in the same time. The distributed treatment order is essential for a regular growth of image objects over the whole scene (Fig. 1). Scanning algorithms or random sequences lead to more or less uneven distributions of image object size over the image.

The *degree of fitting* described in (4) and (5) in connection with the optimization procedure described in the previous sections already yield to non-bad segmentation results. However, they are sub-optimal. They are not based on a clear definition of heterogeneity per image object, which is required for obtaining comparable and optimizable results. Furthermore the criteria are not continuously increasing with the number of merges respectively with the image object scale and therefore may produce image objects with large differences in scale.
Definitions (6) to (9) are certainly more precise. They prevent for instance a homogenous image object to be merged with a heterogeneous image object even if their mean values are similar (Fig. 2).
In contrast to (6) and (7), definitions (8) and (9) are showing continuously increasing values with increasing number of merges resulting in a constant and distributed growth of image objects. A clear definition of heterogeneity per object allows to compare different segmentation results and different optimization procedures.

Especially definition (9) in connection with spectral variance or spectral standard deviation for heterogeneity leads to convincing results. The resulting image objects are highly homogenous, contrasts are represented by image objects consistently and segmentation results are highly reproducible with only slight differences of the edges between objects of low contrast.

Fig. 2: Multiresolution segmentation of airborne image data: trees and meadows are of similar spectral value. They are separated because of the high difference in spectral homogeneity. Following criterion (9), the merge of a homogeneous image object to a heterogeneous image object would result in a big change of heterogeneity for the homogeneous object.

For principle reasons it only makes sense to use the form criteria (12) and (13) mixed with the spectral criteria (10) or (11). Without taking into account spectral information of an image there are no meaningful objects on which form could be optimized. The behavior of the form criteria in terms of optimization is not as good as that of the spectral criteria. Segmentation results are therefore not as reproducible as with spectral criterion only. However, especially in strongly textured image data the form criteria can help to maintain smooth edges or a more or less compact form. Qualitatively, such objects are often much more convincing and useful then objects of an irregular or branched shape.

8 Results

The multiresolution segmentation produces highly homogeneous image objects in arbitrary resolution on different types of data (Fig. 2,3). All three aspects are of considerable advantage and allow to apply this segmentation technique with good results to many different types of image data and problems.

Texture segmentation as also threshold procedures are fully reproducible because they do not rely on criteria which are calculated during the process itself. They are basically using existing criteria or criteria calculated a priori. However, often they don't tre at local contrasts in a satisfying way, quality may vary between image objects. In contrast, the multiresolution segmentation presented here is object oriented: each decision for a merge is based on the concrete recent attributes of the image objects merged in previous steps. This turns out to be the decisive methodical advantage of the method: each decision is based on attributes of homogeneous structures of a recent scale. There are for instance no filter operations typically defined on a (small) fixed-sized window and thereby unaware to generate representative homogeneous areas.

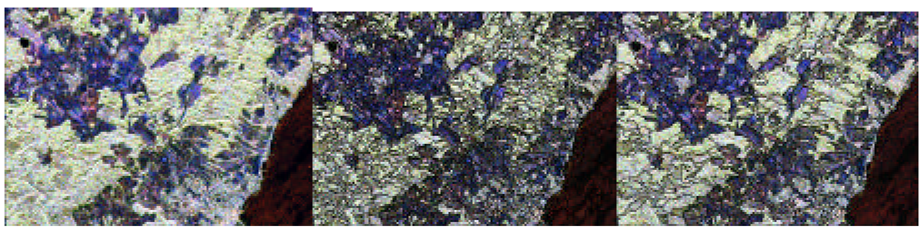

Fig. 3: Image objects in different resolutions extracted based on airborne radar data. This example also demonstrates the capability of multiresolution segmentation to work on relatively strong textured image data.

The advantage of the object-oriented procedure is paid with a certain moment of historicity resulting in the loss of full reproducibility. Different decisions in the past may cause different decisions in the future. However, the arising differences are found on edges between objects of low contrast, where indeed the shape of an edge is more or less arbitrary. If the resulting objects are further classified, in many cases such objects of low contrast are assigned to the same class. Furthermore image objects of smaller scale can be produced instead. On smaller scale fine contrasts are better represented then on larger scale. The differences between the segmentation results therefore will be moved to a smaller scale respectively to finer contrasts, where they do play a less significant role.

Multiresolution segmentation is one basic procedure in the software *eCognition* for object oriented image analysis (Delphi2 Creative Technologies, Munich). It is used here to produce image object primitives as a first step for a further classification and other processing procedures. As a part of the beta-version of eCognition multiresolution segmentation was successfully applied to many different problems in the fields of remote sensing, medical image analysis and structure analysis [BAATZ & SCHÄPE 1999a,1999b; BUCK & al. 1999a, 1999b; DE KOK & al.: 1998, 1999a,1999b; KNOKE 1999; NIEMAYER 1999 & 2000; SCHMIDT 2000, SCHNEIDER & al. 1998].

9 References

Baatz, M. & A. Schäpe (1999a): *GIS-Aktualisierung und Image Fusion basierend auf einem neuen Verfahren zur objektorientierten und multiskalaren Bildanalyse*. In: Strobl, J. & T. Blaschke (Hrsg): Angewandte Geographische Informationsverarbeitung XI. Beiträge zum AGIT-Symposium Salzburg 1999. Karlsruhe. Herbert Wichmann Verlag.

Baatz, M. & A. Schäpe (1999b*): Object-Oriented and Multi-Scale Image Analysis in Semantic Networks*. In: Proc. of the 2nd International Symposium on Operationalization of Remote Sensing August 16th – 20th 1999. Enschede. ITC.

Bovik, A.; Clark, M. & W. Geisler (1990): *Multichannel texture analysis using localized spatial filters*. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 12.

Buck, A.; de Kok, R.; Schneider, Th. & U. Ammer (1999a): *Improvement of a forest GIS by integration of remote sensing data for the observation and inventory of "protective forests" in the Bavarian Alps*. In: procc. IUFRO Conference on Remote Sensing and Forest Monitoring, June 1-3,1999, Rogow, Poland.

Buck, A.; de Kok, R.; Schneider, Th. & U. Ammer (1999b): *Integration von Fernerkundung und GIS zur Beobachtung und Inventur von Schutzwäldern in den Bayerischen Alpen*. In: AGIT 99, "Angewandte Geographische Informationsverarbeitung XI", Hrsg. Strobl/Blaschke, ISBN 3-87907-336-8, S 94-101.

Chaudhuri, B. & N. Sarkar (1995): Texture segmentation using fractal dimension. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 17, Nr. 1, S.72-77.

Derin, H & W. Cole (1986): *Segmentation of textured images using Gibbs random fields*. In: Computer, Vision, Graphics, and Image Processing. Vol. 35, Nr. 1, S. 72-98.

Geman, D.; Geman, S.; Graffigne, C & Dong, P. (1990): *Boundary detection by constrained optimization*. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, Nr. 7, S.609-628.

Gorte, B. (1998): *Probabilistic Segmentation of Remotely Sensed Images*. In: ITC Publication Series No. 63.

Haralick, R.; Shanmugan, K & I. Dinstein (1973): *Textural features for image classification*. In: IEEE Transactions on Systems, Man and Cybernetics. Vol. 3, Nr. 1, S.610-621.

Hofmann, T.; Puzicha, J. & J. Buhmann (1998*): Unsupervised texture segmentation in a deterministic annealing framework*. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 20, Nr. 8, S. 803-818.

Jain, A. & F. Farrokhnia (1991): *Unsupervised texture segmentation using Gabor filters*. In: Pattern Recognition. Vol. 24, Nr. 12, S. 1167-1186.

deKok, R; Schneider, Th. & U. Ammer (1998): *Das Problem der Schatthänge im Luftbild - können digitale CCD Scanner die Informationslücke schließen?*. In: AFZ 24, S. 1454-1457.

deKok, R; Schneider, Th. & U. Ammer (1999a*): Object based classification and applications in the Alpine forest environment*. In: procc. Joint ISPRS/EARSeL Workshop "Fusion of sensor data, knowledge sources and algorithms, Valladolid, Spain, June 3-4, 1999.

deKok, R.; Schneider, T.; Baatz, M. & U. Ammer (1999b*): Object based image analysis of high resolution data in the alpine forest area*. In: Joint WSf ISPRS WG I/1, I/3 and IV/4: SENSORS AND MAPPINHG FROM SPACE 1999; Hannover, September 27-30, 1999.

Knoke, R. (1999): *Extraktion objektstrukturierter Geoinformationen aus Fernerkundungsdaten*. Unveröffentlichte Diplomarbeit. Universität der Bundeswehr München.

Laine, A. & J. Fan (1996): *Frame representations for texture segmentation*. In: IEEE Transactions on Image Processing. Vol. 5, Nr. 5, S.771-779.

Manjunath, B. & R Chellappa (1991): *Unsupervised texture segmentation using Markov random field models*. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 13, S. 478-482.

Mao, J & A. Jain (1992*): Texture classification and segmentation using multiresolution simultaneous autoregressive models*. In: Pattern Recognition, Vol. 25, S. 173-188.

Niemeyer, I. (1999): *Fractal-hierarchical Pattern Recognition for Safeguard Purposes*. In: Proc. of the 2nd International Symposium on Operationalization of Remote Sensing August 16th – 20th 1999. Enschede. ITC.

Niemeyer, I. et al, (2000): *Object-Based and Multi-Scale Analysis of Satellite Imagery Using Semantic Nets for Safeguard Purposes*. In: Fourth International FLINS conference on intelligent Techniques and Soft Computing in Nuclear Science and Engineering, Bruges, Belgium August 28th – 30th 2000.

Panjwani, D. & G. Healey (1995*): Markov random field models for unsupervised segmentation of textured color images*. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 17, Nr. 10, S. 939-954.

Puzicha, J. & J.M. Buhmann (1998): *Multiscale Annealing for real-time unsuperviesed Texture Segmentation*. Technical Report IAI-97-4; Rheinische Friedrich-Wilhelms-Universität.

Salari, E. & Z. Ling (1995): *Texture Segmentation using hierarchical Wavelet Decomposition*. In: Pattern Recognition, Vol.28, Nr.12, S. 1819-1824.

Schmidt, R (2000): *Untersuchung des Bildanalysesystems eCognition*. Unveröffentlichte Diplomarbeit; Universität Hannover.

Schneider, Th.; Buck, A.; de Kok, R. & W. Semmt (1998): *Konzepte, Probleme und erste Ergebnisse eines auf Fernerkundung gestützten Monitoring-Systems für den Alpenraum*. Forstliche Hochschulwoche, Freising.

Shi, J. & J. Malik (1997): *Normalized cuts and image segmentation*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR ´97), S. 731-737.

Wegner, S., et al., 1997, Segmentierung mit der Wasserscheidentransformation; Spektrum der Wissenschaft 6/1997, pp. 113 - 115

Won, C.S. & H. Derin (1992): *Unsupervised segmentation of noisy and textured images using Markov random fields*. In: CVGIP: Graphical Models and Image Processing. Vol. 54, S.308-328.