Journal of Microscopy

Journal of Microscopy, Vol. 250, Pt 2 2013, pp. 130–141
Received 4 July 2013; accepted 20 February 2013

doi: 10.1111/jmi.12029

Extraction and parametrization of grain boundary networks in glacier ice, using a dedicated method of automatic image analysis

T. BINDER*, C.S. GARBE*, D. WAGENBACH†, J. FREITAG‡ & S. KIPFSTUHL‡

**Interdisciplinary Center for Scientific Computing, Universität Heidelberg, Heidelberg, Germany*

†Institut für Umweltphysik, Universität Heidelberg, Heidelberg, Germany

‡Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany

Key words. Grain boundary network, ice microstructure, image segmentation, machine learning, surface sublimation etching.

Summary

Microstructure analysis of polar ice cores is vital to understand the processes controlling the flow of polar ice on the microscale. This paper presents an automatic image processing framework for extraction and parametrization of grain boundary networks from images of the NEEM deep ice core. As cross-section images are acquired using controlled surface sublimation, grain boundaries and air inclusions appear dark, whereas the inside of grains appears grey. The initial segmentation step of the software is to separate possible boundaries of grains and air inclusions from background. A Machine learning approach is utilized to gain automatic, reliable classification, which is required for processing large data sets along deep ice cores. The second step is to compose the perimeter of section profiles of grains by planar sections of the grain surface between triple points. Ultimately, grain areas, grain boundaries and triple junctions of the later are diversely parametrized. High resolution is achieved, so that small grain sizes and local curvatures of grain boundaries can systematically be investigated.

Introduction

Deep polar ice cores drilled during the last decades are not only valuable archives of the past climate, they also sample a forever ongoing natural deformation and flow experiment. Ice deforms and flows under the overburden accumulated snow and ice. Microstructure analysis along ice cores of up to more than 3000-m length yields experimental data on the evolution of the microstructure through ice sheets and the processes controlling ice flow like deformation and recrystallization. It has been shown before that strain induced recrystallization occurs in metals (e.g. Beck & Sperry, 1950; Humphreys & Hatherly, 2004) and ceramics (e.g. Heuer *et al.*, 1969; Mueller, 1935). Knowledge of these processes is important as they determine the material properties of flowing ice on geological time scales difficult to reproduce under laboratory conditions. A precise understanding of ice sheet flow is required not only for ice core dating or reconstructing the past climate, but also to predict the climate in the future. In the expected warmer world, enhanced ice flow leads to a rising sea level. Ice flow simulations are based on empirical material laws. Ice is a highly anisotropic material. Although one of the purest natural materials on the Earth's surface, ice in the ice sheets is not at all homogeneous. Impurities on trace levels, varying on seasonal and climatic cycles, affect the flow properties of ice on all scales. This may cause the inhomogeneous deformation and recrystallization phenomena which can enhance the ice flow rate by a factor of three or more (Budd & Jacka, 1989). Microscopic processes control both the flow of ice and the development of physical properties such as grain size and *c*-axis orientation. From observations of the physical properties the microscopic processes operating can be inferred and the laws governing the flow be estimated (Alley, 1992). A comprehensive analysis describing the microstructure along ice cores by different physical properties provides a better understanding of the flow of polar ice sheets and glaciers.

The method of optical microscopy used here utilizes controlled sublimation (Arnaud *et al.*, 1998; Kipfstuhl *et al.*, 2006). The state of the art in analysis of sublimation images is a semimanual extraction, necessitating interaction for each image. Precise information about the microstructure within an ice sheet and thus the ice flow can be gained from parallel sections with small distances along the vertical direction. As the extracted cross-sections of grains are of the order of millimetre to centimetre, sections (taken at a resolution of 5 μm) need sufficient size to be representative. Therefore, a large, quasi-continuous data set has to be processed automatically and efficiently with just a few steps of manual adaption

Correspondence to: T. Binder, Interdisciplinary Center for Scientific Computing, Universität Heidelberg, Speyerer Straße 6, D-69115 Heidelberg, Germany. Tel: +49 6221 54 8103; fax: +49 6221 54 8790; e-mail: Tobias.Binder@iwr.uni-heidelberg.de

(creation of training data and selection of filter parameters) at the beginning. For example, for the NEEM ice core about 1000 sections have to be analysed. In the following, it is assumed that the two-dimensional cross-sectional area of a grain (derived from a section of an ice sample) is a measure of the true three-dimensional grain size and is called grain size.

In this contribution, a new method to extract and parametrize grain boundaries and entire networks of grain boundaries from images of ice core sections acquired at microscopic resolution is introduced. Compared to previous work this technique extends the quantitative description of the evolution of the microstructure along ice cores and through ice sheets, respectively. It is shown exemplary that the technique using controlled sublimation allows for a more detailed description of ice microstructure and its variability within a section and through an ice sheet than before. Focus is laid on the extraction of:

- small grains (equivalent radius of 50 μm),
- local curvature of grain boundaries and
- angles in grain triple junctions.

Methods

Denotations

For simplification, section profiles of grains are referred to as 'grain areas' and planar sections of the grain surface are referred to as 'grain boundaries'. They are visible as black lines in an image of a carefully polished section of ice as described in detail by Kipfstuhl *et al.* (2006). A (single) grain boundary is terminated by two triple junctions (as long as it is not the open end at the margin of a section). Consequently, the term triple point/junction corresponds to planar sections, angles are denoted as planar angles.

Image acquisition

Polarized light microscopy of thin sections is well-established in structural geology and in microstructure analysis of ice. In this work, sublimation patterns on the surface of carefully polished ice core sections are mapped in reflected (unpolarized) light mode at a resolution of 5 μm. This technique does not reveal crystal orientation but documents subgrain boundaries and, therefore, areas of low-angle lattice distortions (Weikusat *et al.*, 2011). As the grain boundaries in this method are imaged directly, the smallest grain size is in principle only limited by the width of a grain boundary. This is contrasted by the smallest meaningful size in a thin section between crossed polarizers which depends on the sample thickness.

The images used in this work for exemplary results stem from the NEEM ice core (North–West Greenland, 2537 m length) drilled in 2008–2010. Vertical ice core sections are about 4 × 9 cm^2 large and are prepared as described by Kipfstuhl *et al.* (2006). In contrast to the latter, the field of view is imaged in a single step by a Large Area Scanning Macroscope (Schäfter & Kirchhoff GmbH) assembling the image line by line with 8192 pixels. It utilizes directed bright-field illumination, light is radiated from the direction of the camera. Areas oriented parallel to the sensor reflect much light into the camera and appear bright. Structured areas and skewed edges appear dark.

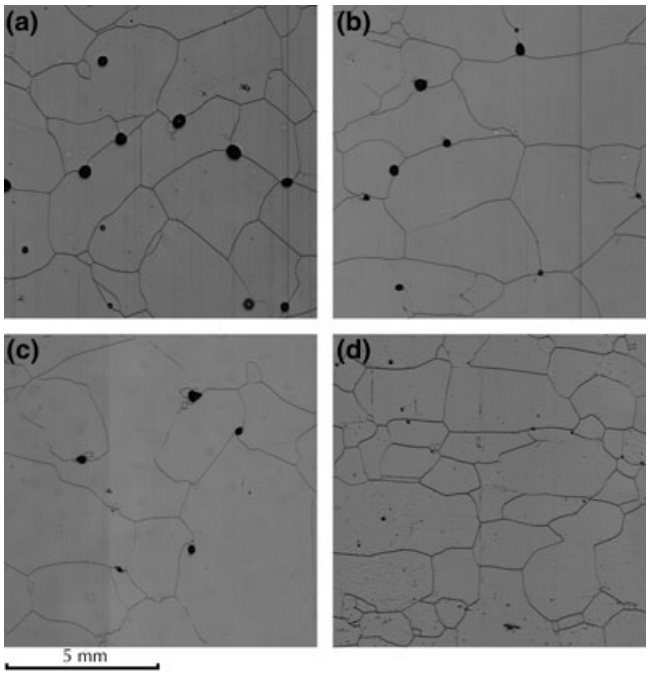

Fig. 1. Image sections from different depths: (a) 213 m, (b) 361 m, (c) 581 m, (d) 1499 m.

In Figure 1, image sections from four different depths are shown. Air inclusions and most grain boundaries (sublimation grooves) appear dark, whereas the inside of grains appears grey. The task for image analysis is to extract predominantly dark grain boundaries and such of air inclusions reliably.

Image processing

Comparison to previous work

Arnaud *et al.* (1998) were the first who analysed sublimation images by means of image processing. However, this work has been restricted to the extraction of grain sizes in firn. A comprehensive parametrization of grain boundary networks contains different structural, morphological and topological parameters, as has occasionally been done for images acquired by polarized light microscopy (e.g. Gay & Weiss, 1999). A single grain boundary forms the basic element for parametrization of the grain boundary network. The method presented here utilizes up-to-date machine learning techniques to exploit grey value profiles across mapped grain boundaries and manually generated training data to the extent feasible. The error rate of

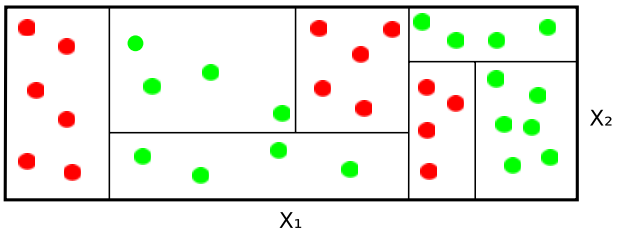

Fig. 2. Completely separated classes in two-dimensional feature space, red and green points represent training points in two classes.

classification can iteratively be reduced by adding new filters or training data. In contrast to this, the techniques utilized by Arnaud *et al.* (1998) are much more static and tuned to data. The same applies to the image analysis program Object Image (Vischer *et al.*, 1994) used by Kipfstuhl *et al.* (2009).

An automatic software approach for extraction and parametrization of the section profiles of ellipse-shaped air bubbles has already been developed (Ueltzhöffer *et al.*, 2010).

Random forest classification

Supervised Machine Learning (e.g. Hastie *et al.*, 2009) is applied for image classification. A mapping from the usually high-dimensional feature space to a discrete set of classes is created based on manually generated training data. Initially, pixels as basic classification elements are mapped to classes 'boundary' and 'no boundary'. Subsequently, boundaries are mapped to 'grain boundary', 'bubble boundary' or 'artefact'. Both steps are implemented by Random Forest classifiers (Breiman, 2001). It has been shown that this type of approach exhibits highly accurate results (Ren & Malik, 2003). A classifier consists of randomly created decision trees which are Embryo zebrafish segmentation using an improved hybrid method grown as follows:

(1) If N feature vectors have been manually classified, choose repeatedly one of them at random but with replacement until a sample of N training points (called bootstrap sample) is created. Some training points may be contained several times whereas others do not appear at all.
(2) Choose randomly m features (with $m \ll M$ and M being the number of features) which are used to divide the feature space into two subsets. Different 'impurity measures' may identify the 'best split' (i.e. a threshold for every feature coordinate X_i with $i \in [1..m]$) that separates in the best case two subsets containing only training points of one class. The value of m is kept constant during the forest growing.
(3) Each tree is grown until classes have completely been separated (Fig. 2).

One decision tree may be very noisy and prone to overfitting, but considering a high number (in the present case 256) and averaging over their 'votes' may yield more accurate probabilities of belonging to a certain class.

The dimension of feature space should be as high as necessary to allow effective classification. Since every additional feature compromises the efficiency, a measure for feature variable importance is needed. In the example of Figure 2 four decisions are made on variable X_1 and two decisions are made on variable X_2. Omitting a feature variable where many decisions are based on may corrupt the random forest prediction more than omitting a feature barely used for decisions. Mean Decrease Gini (MDG) is directly correlated with the sum of decisions made on a certain feature over all trees in the forest and allows for robust random forest feature ranks (Calle & Urrea, 2010). Additionally, the detection of critical boundaries has been verified for the selection of features.

Applied filters

The pixel feature space is constructed by the responses of linear and shift-equivariant filters, which are based on the scalar, vectorial and tensorial quantities described in the following. Some frequently used filters can be expressed by the convolution of an image with the Gaussian kernel k_σ (standard deviation σ). In the present context, k_σ is approximated by the discrete analog of the Gaussian introduced by Lindeberg (1994) with radius

$$r = \sigma \times (3 + 0.5 \times n), \text{ where } n \text{ is the order of the derivative filter.} \quad (1)$$

The second term accounts for the fact that the derivative of the Gaussian becomes wider with increasing order, the radius is rounded to the next higher integer.

The first order derivative filter $\vec{g}(f)$ is given by the gradient of the two-dimensional Gaussian kernel with the parameter σ. Accordingly, the Laplace operator leads to the second order Laplacian of Gaussian filter.

The derivative operator can also be given by the entries of the Hessian matrix

$$H = \begin{pmatrix} \frac{\partial^2}{\partial x \partial x} & \frac{\partial^2}{\partial x \partial y} \\ \frac{\partial^2}{\partial y \partial x} & \frac{\partial^2}{\partial y \partial y} \end{pmatrix}. \quad (2)$$

Diagonalization of the resulting tensor yields the eigenvalues λ_1, $\lambda_2 (\lambda_1 \geq \lambda_2 \geq 0)$ and the corresponding eigenvectors $\vec{e}_1$, $\vec{e}_2$ – the principal directions of pure curvature.

Information on the predominant directions of the gradient is given by the eigenvalues λ_1, λ_2 resp. the eigenvectors $\vec{e}_1$, $\vec{e}_2$ of the Structure Tensor (Förstner, 1986; Harris & Stevens, 1988). The basis is provided by the operator Q generating the

gradient tensor of an image (Köthe, 2003a):

$$Q = (g_x, g_y)^T \cdot (g_x, g_y). \tag{3}$$

A second Gaussian kernel with standard deviation σ' defines a window for spatial averaging. The operator generating the structure tensor has the form

$$S_{\sigma,\sigma'} = (s_{ij}), \qquad s_{ij} = k_{\sigma'} \star q_{ij} \qquad (i, j \in \{1, 2\}), \tag{4}$$

where $\star$ denotes convolution.

The Boundary Tensor introduced by Köthe (2003b) is defined by a quadrature filter approach as the sum of an odd and an even tensor. The odd tensor is given by a generalized form of the structure tensor, whereas the even tensor is generated by a generalized Hessian.

$$T^{(\text{Boundary})} = T^{(\text{odd})} + \frac{1}{4} T^{(\text{even})} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \tag{5}$$

$$E_{\text{Boundary}} = \text{tr}(T^{(\text{Boundary})}) = b_{11} + b_{22} \tag{6}$$

Irrespective of the symmetry of the edge (roof or step edge), the boundary energy represented by the trace of the boundary tensor has a single peak at the edge position.

Information on boundary direction can also be extracted from the boundary tensor as the structure tensor and the Hessian matrix inform how the gradient and least curvatures are predominantly oriented.

Image-processing framework

The framework is divided into three steps: segmentation, microstructure extraction and microstructure parametrization. It has been implemented as plain C++ code using open source libraries/software packages. Random forest classification and watershed region growing (segmentation algorithm) are implemented by means of Vigra (hci.iwr.uni-heidelberg.de/vigra/). Even plotting and a Graphical User Interface have been embedded in the framework using PLplot (plplot.sourceforge.net), CImg (cimg.sourceforge.net) and Qt (qt.digia.com). The advantage of this framework is full independence of any commercial software as well as high performance.

Segmentation

In Figure 3, an overview on performed steps of segmentation is given. To exploit the potential of modern multicore computers, time-consuming algorithms (for calculation of filters, random forest prediction and parametrization of segmentation) have been implemented parallelizable utilizing OpenMP (openmp.org) and Boost MPI (boost.org/libs/mpi).

The watershed region growing algorithm (Vincent & Soille, 1991) is very appropriate as it yields areas corresponding to grains closed up by adjacent grain boundaries.

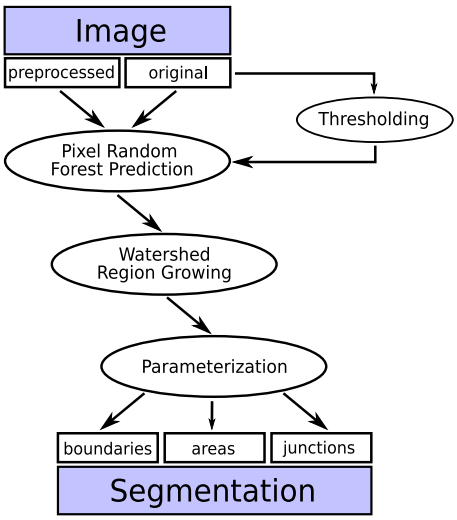

Fig. 3. Overview from image to segmentation.

For preprocessing, a Laplace of Gaussian as second-order derivative filter with standard deviation $\sigma = 2.0$ is applied to highlight the centre of grain boundaries (roof edges) as local maxima. Bubble boundaries (step edges) lead to local minima and maxima close together. The detection of connected grain boundaries in the image can be complicated by artefacts, such as reflections, or by boundaries becoming lighter in the vicinity of a bubble. To overcome this, a morphological minimum operator with a disc of radius 5 pixels is applied to the filtered image. In Figure 4, the result of the Laplace of Gaussian filter and closure of small gaps (up to 5 pixels) is shown. Even though a higher disc radius would close larger gaps in the image, an application is not reasonable, since structures smaller than the disc radius would vanish. High gradients resulting from the applied minimum operator are smoothed by a Gaussian kernel with $\sigma = 0.1$.

Applying watershed region growing to the result of the pixel random forest classifier, a probability map, yields a low ratio of artificial segmentation boundaries. The degree of oversegmentation is higher if applied to the preprocessed image (see Fig. 5). Both critical grain boundaries (1 and 2) marked in Figure 4(a) occur among the segmentation boundaries.

Bubble boundaries in the probability map can contain discontinuities (see Fig. 5), giving point to an additional regularization step, otherwise some bubble boundaries would vanish because of penetrating watersheds. This anisotropic smoothing is performed with the objective of preserving thin structures, the method itself has been proposed by Tschumperlé (2006).

Training data

Effective random forest classifiers are generated if the training data contain the whole variety of structures in a balanced and

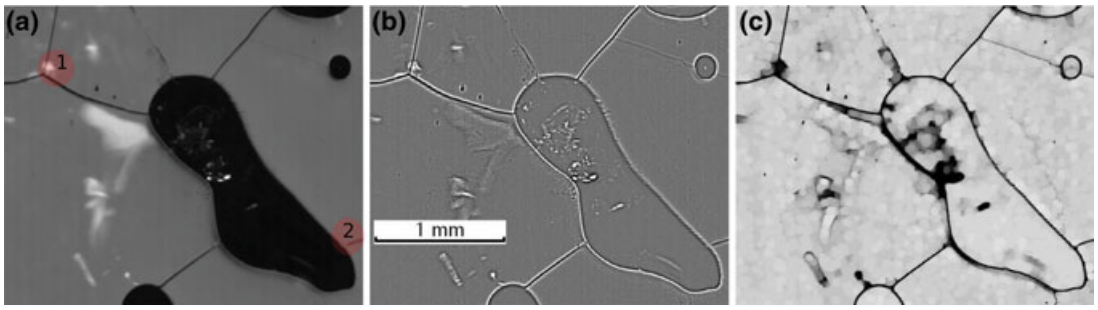

Fig. 4. Different steps of preprocessing: (a) original grey value image: (1) reflection beneath the grain boundary, (2) grain boundary becoming light in the vicinity of a bubble. (b) image after Laplace of Gaussian filter. (c) image after application of minimum operator – (NEEM, 58 m).

Fig. 5. (a) Segmentation based on preprocessed image. (b) Probability map. (c) Segmentation based on probability map.

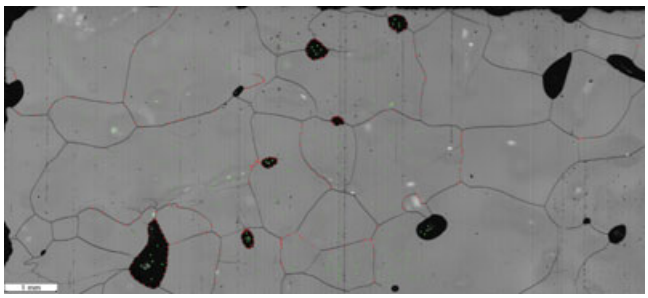

Fig. 6. Training data containing about 400 pixels (200 red points for class 'boundary' and 200 green points for class 'no boundary') in a section from 102 m depth.

redundant way. Otherwise, the classifier depends significantly on the bootstrap samples found at random. Shortcomings of a classifier can be overcome by additional training data. The present results (images from different depths) are based on one training data set of about 400 pixels set in a section from shallow depth (Fig. 6), which is generally applicable. With ascending number of pixels, the robustness of the classifier increases. However, 400 pixels is a reasonable number as the enhancement above 300 pixels becomes negligible.

Features

The set of features (filters) utilized for pixel classification can be optimized iteratively. As it can be seen in Figure 1, the grey values of grain boundaries show significant differences compared within a single section and with each other. Features based on the grey value do not allow to segment grain boundaries reliably and to keep the ratio of artificial segmentation boundaries low. Currently, a five-dimensional feature space is used which represents a good compromise between robustness and computational effort:

- first eigenvalue of the structure tensor of the preprocessed image ($\sigma = 1.0, \sigma' = 0.6$),
- first eigenvalue of the structure tensor of the preprocessed image ($\sigma = 1.5, \sigma' = 0.9$),
- second eigenvalue of the Hessian matrix of the preprocessed image ($\sigma = 2.0$),
- trace of the boundary tensor of the preprocessed image representing the boundary energy, see Eq. (6),
- difference of two Gaussian smoothed original images ($\sigma_1 = 0.8$, $\sigma_2 = 6.0$): Gaussian smoothed step boundaries differ substantially depending on the chosen σ. The margin between two differently smoothed images is utilized to highlight the centre of bubble boundaries.

For the structure tensor, the optimal choice of parameters depends strongly on the width of the structures to be extracted. As the width of boundaries is between 3 and 8 pixels two different settings are used. For the Hessian matrix filter, one optimal choice has been found. In case of too low parameters blurred structures will not be segmented. At the other extreme,

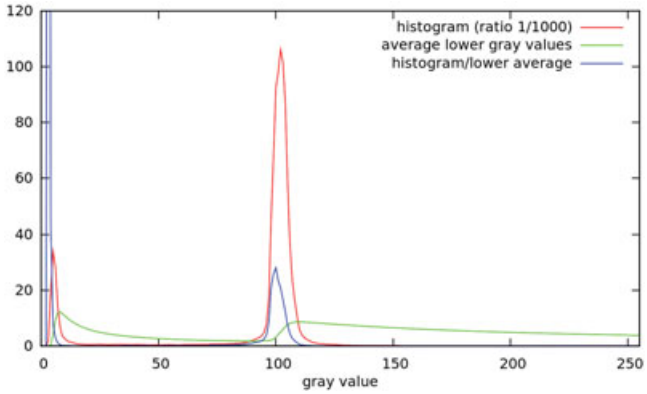

Fig. 7. Frequency distribution of grey values (red), average of frequency values $[0..x-1]$ for grey value x (green), ratio between frequency value and lower average (blue). The latter is used for thresholding.

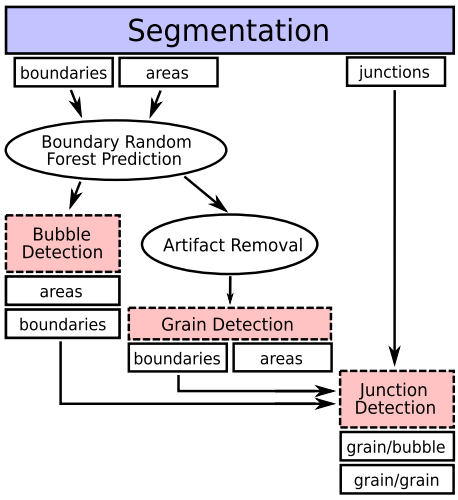

Fig. 8. Overview from segmentation to grain boundary network.

resolution in the position of the centre of boundaries is lost. The NEEM data set has been acquired at approximately constant conditions (same lens system), the choice of parameters can be applied to all images.

Detection of relevant regions

The random forest prediction forms the most time-consuming part of the segmentation step. Since large data sets from deep ice cores are intended to be processed, optimization techniques besides parallelization need to be developed. By reducing the random forest classification to relevant regions, a significantly better performance can be achieved.

Irrelevant pixels are rated with low-boundary probability. Since the selection of seeds for the watershed region growing algorithm is critical (Köthe, 1995), these pixels are additionally used as seeds. Otherwise, it may happen casually that smaller grains next to bubbles are not reached by a watershed, leading to a missing bubble boundary in the segmentation. If every grain contains detected irrelevant pixels in the inside, a watershed starts to grow in every grain.

Pixels for which classification can be skipped are found based on their grey value since grain boundaries and bubbles are much darker than the grain's inside. Robust results are achieved by a grey value threshold adapted to the grey value frequency distribution of the whole image (Fig. 7). The shape of this distribution is very similar for all processed images: beside a peak at very low grey values (very dark in the image) due to bubbles and grain boundaries, there is a second peak at higher grey values, representing most pixels in the grains' inside. The level of other grey values is quite low, but differs significantly from image to image.

The heuristic approach implemented for the calculation of grey value thresholds involves the comparison between the frequency of a given grey value and the average of frequencies of lower grey values. Ultimately, the ratio between the 'actual' grey value and the lower average is taken. The higher the frequency of the threshold lies above the frequencies of lower grey values, the more pixels at the transition between the grains' inside and the grain boundaries are preserved. These pixels seem to be critical and thresholding of the ratio pointed out to be easier and more robust, than thresholding of the derivative of the frequency. Regions of very high brightness are preserved from thresholding by a second grey value threshold, since these regions may hide grain boundaries (as the reflection marked by 1 in Fig. 4 a).

Boundaries and junctions are affixed by a unique label. The pixels of a boundary are sorted so that they start and end at junctions. This is needed for the calculation of orientation and curvature.

Microstructure extraction

At this stage, relevant information of the original image has been reduced to possible boundaries (segmentation boundaries) which have to be classified into three classes: no-boundary, grain boundary and bubble boundary.

This module makes use of a segmentation with high variability in boundary length and area size, which is in general not given. Mostly, segmentation results look like in Figure 5(a). In the present case, longer segmentation boundaries are most likely part of the grain boundary network and larger areas correspond to grains with high probability. In Figure 8, an overview on the model developed for the extraction of the grain boundary network is given.

Boundary random forest classification

About 300 boundaries among the boundary candidates of different images have been classified manually. For all

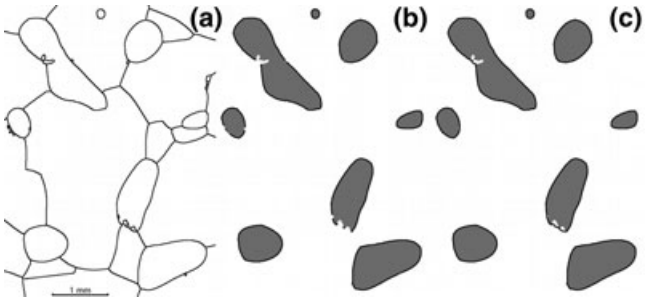

Fig. 9. Bubble closing using Dijkstra's shortest path algorithm. (a) Segmentation. (b) Bubble areas with selected boundaries before closing. (c) Result of closing (NEEM, 58 m).

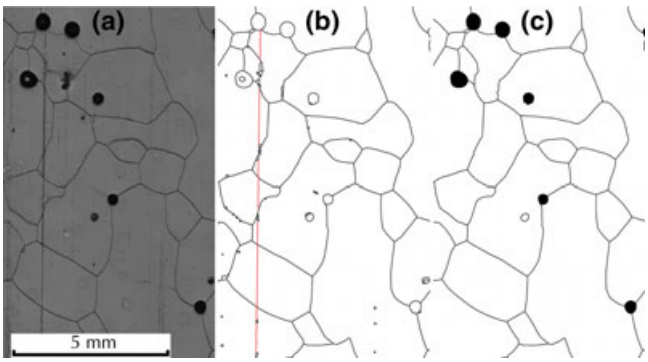

Fig. 10. Artefact removal by Hough transform. (a) Original image. (b) Segmentation, found vertical structures marked red. (c) Final network structure (NEEM, 168 m).

segmentation boundaries a random forest prediction is calculated based on the following four boundary features:

- **Pixel probability mean:** watershed region growing is based on the probability map (the result of the pixel random forest classifier). Most obtained segmentation boundaries are located at maxima of the probability map, consequently the mean of pixel probability along a segmentation boundary is predominantly high. If this is not the case, the boundary is most likely artificial.
- **Difference of mean grey values of adjacent areas:** bubble boundaries can be identified by having a dark area on one side and a light area on the other side.
- **Sum of neighbour areas size:** it is most improbable that two large areas in the segmentation need to be merged to form one even larger grain. In contrast, two small areas can correspond to two small grains just as well, as they need to be merged. This boundary feature preserves two large areas from being merged.
- **Absolute curvature mean:** the mean absolute curvature along a bubble boundary is generally higher than along a grain boundary. Besides, some artificial boundaries show significantly higher averages of absolute curvature.

If a boundary is classified as artefact, the two adjacent areas will be merged. A function for area merging has been implemented, which is called every time a type of artificial boundaries is found.

Bubble detection

The idea is to find areas in the segmentation with high-bubble probability. This probability is calculated by averaging over the bubble boundary probability of all adjacent boundaries. For areas with bubble probability above a defined threshold, individual bubble boundaries with high probability are selected (Fig. 9). In some cases, bubble boundaries are already closed or can be connected easily as only two or three segmentation boundaries enclose the found area. Otherwise, a loop of bubble boundaries is found in the following way:

(1) find two junctions with maximal distance reached by selected bubble boundaries, which are referred to as A and B.
(2) Weight all selected bubble boundaries with zero, all others by their length in pixels.
(3) Use Dijkstra's algorithm (e.g Cormen *et al.*, 2009) to find a shortest path from A to B. Weight all boundaries used with infinity.
(4) Find a shortest path from B to A using Dijkstra's algorithm.

Artefact removal

As it can be seen in Figure 8, only areas and boundaries of bubbles are found so far. Prior to the detection of grains, two different types of artificial boundaries are handled in the step denoted 'Artefact Removal': vertical structures caused by microtoming (see Fig. 10) and boundaries artificially created by watershed region growing.

Microtoming the ice surface is done very carefully and only in vertical direction. Thus, artificial structures can be identified by being straight and inclined to the vertical by less than 2°. If they were not inclined at all, it would have been possible to create a histogram of grain boundary pixels along the horizontal direction and to find maxima. To take the inclination into account a two-dimensional Hough space is created instead of an one-dimensional histogram (as every straight line can be described by a set of two parameters).

Since grain boundaries and scratches caused by microtoming are mapped similar by the camera, it must be ensured that only artificial segmentation boundaries are removed. The applied criterion utilizes the arrangement of pixels to boundaries and cannot be used at an earlier stage. Artificial segmentation boundaries caused by microtoming:

- start and end on a straight line parametrized by a maximum in Hough space,
- show small deviation in horizontal direction, absolutely as well as compared to the vertical deviation along the whole boundary.

In order to find boundaries artificially created by watershed region growing at sites where the pixel probability is low, the boundary feature 'Pixel probability mean' has been added. As segmentation boundaries of this type are part the boundary training data set, the boundary random forest classifier assigns high probability for class 'artefact' to similar ones. Segmentation boundaries exceeding a threshold (of 0.9 or higher) are removed.

Bubbles and different types of artefacts have been identified, so that all remaining unclassified areas and boundaries can be interpreted as grains resp. grain boundaries (Fig. 10).

Grain and junction detection

Even though small grains can form in the ice sheet, the smallest areas are most probably artefacts of the segmentation algorithm. Examples for small artificial areas can be seen next to the bubble boundary in Figure 5(c) or in the inside of grains in Figure 10(b). The easiest way to identify physical meaningful areas as grains is to threshold of their size. Small grains cover only a small volume fraction, but a grain size distribution weighted by the number will be more meaningful if small grains are included to the greatest possible extent. Choosing a grain size threshold has to be well balanced between excluding as few small grains and as many artefacts as possible.

Identifying areas as artefact necessitates the selection of a neighbouring area to merge with, resp., the selection of a boundary separating both areas. It has been tested to merge artificial areas:

- along the boundary with lowest boundary probability,
- along the boundary with highest mean curvature,
- along the longest boundary (if two areas share several boundaries, the length of all boundaries is taken).

The first two criteria may be the best choice in some special cases, but the last one yields the most robust results.

The function for grain merging is called iteratively since erroneous merging results can occur, if one grain is merged with several neighbours during one merging step. Therefore, each grain can be merged only once, after 2–4 iterations all artefacts have been removed. The physical motivation for this merging is the conservation of mass. If artificial structures inside a grain are removed, their area will be added to the surrounding grains leaving the total area unchanged.

Identified bubble and grain boundaries can be combined to grain–grain junctions and grain–bubble junctions: If two bubble boundaries and at least one grain boundary meet, the junction is classified as grain–bubble junction. If only grain boundaries meet the junction is denoted as grain–grain junction.

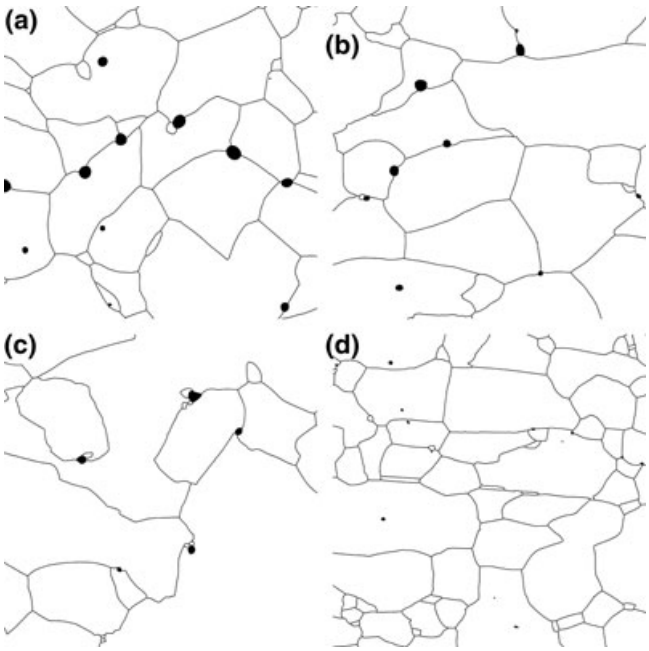

Fig. 11. Extracted grain boundary network for the image sections shown in Figure 1.

Results

Using the image sections shown in Figure 1, a task for the image analysis approach has been defined. For the results depicted in Figure 11, a grain size threshold of 500 pixels (corresponding to an equivalent grain radius of 65 μm) is used. This choice allows for a reliable automatic extraction. For even smaller grain sizes interactive extraction is feasible.

Microstructure parametrization

Grain areas. Based on the previous step, it has become possible to derive different parameters, from which a selection is described here. A variety of parameters is related to grain areas and bordering grain boundaries (Fig. 12):

- grain size: equivalent radius (radius of a circle with the same area),
- number of grain boundaries/grain neighbours,
- grain width w and height h (see box in Fig. 12 a),
- aspect ratio in vertical direction (w/h),
- length of longest grain boundary (see Fig. 12 a),
- roundness (grain area size over square of grain perimeter),
- long axis angle of fitted ellipse,
- aspect ratio of fitted ellipse (length of long axis over length of short axis),

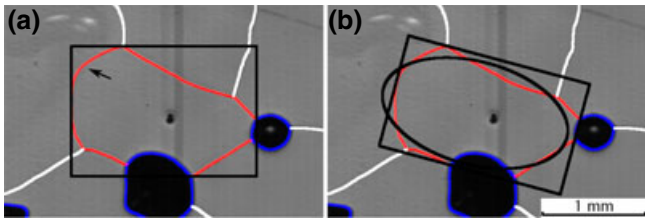

Fig. 12. Parametrization of grain areas. (a) Width and height determined by outer box, longest grain boundary marked by arrow. (b) Direction of largest elongation found by fitted ellipse, outer box along long axis of the ellipse defines 'inclined' width and height (NEEM, 102 m).

- aspect ratio of an outer box oriented along the ellipse axes (see box in Fig. 12 b).

Although the last two parameters are intimately related, they are not equivalent. For 5000 grains from shallow depths, a Pearson's correlation coefficient for the last two parameters of $\rho \approx 0.98$ is calculated.

Grain boundaries

Grain boundaries are parametrized by their length and pixelwise estimated curvature. The calculation of continuous curvature values from discrete pixel positions of the grain boundaries necessitates different smoothing and interpolation techniques. The following implementation has been chosen to meet this requirements:

- reduce the resolution (number of pixels) of boundaries to the half by replacing pairs of pixel positions by one mean.
- calculate the components of the local tangent vector (at the ith pixel) Δx_i and Δy_i using a discrete 5 pixels approximation (Richardson differentiation scheme): $\Delta x_i = \frac{1}{12}(x_{i+2} - x_{i-2} + 8x_{i+1} - 8x_{i-1})$, analogously for Δy_i.
- despite the initial averaging and the approximation for Δx and Δy, an additional smoothing is necessary since the exact position of the boundary can vary by several pixels, especially for broad and homogeneous black grain boundaries. Therefore, a convolution with binomial mask is applied: $\Delta x'_i = \frac{1}{16}(\Delta x_{i+2} + 4\Delta x_{i+1} + 6\Delta x_i + 4\Delta x_{i-1} + \Delta x_{i-2})$, analogously for $\Delta y'_i$. The factors in this mask are binomial coefficients.
- use Δx and Δy to calculate the angle of the local normal vector along the boundary by $\phi_i = \mathrm{atan2}(\Delta y_i, \Delta x_i)$. atan2 is a variation of the arctangent function defined in a variety of computer languages with the value range $(-\pi, \pi]$.
- if $\Delta x < 0$, atan2 yields 2π flips for changing sign of Δy. Those flips have to be found as they would cause artificially high curvatures. This can be done by comparing ϕ with (up to 5) previous values. If necessary, 2π will be added or subtracted.
- with 2π flips being removed, the orientation angle ϕ can be convolved with a binomial mask to smooth other 'noise'.
- at this stage, the boundaries can be resized to their original length by means of interpolation.
- 'quasi-continuously' curvature values κ_i can be calculated by $\kappa_i = \Delta\phi_i / \Delta s_i$, where $\Delta\phi_i$ is the change of the local normal vector along the boundary length Δs_i from the ith to the $(i+1)$th pixel.
- Finally, the curvature κ is smoothed in the same way as described above for Δx_i (convolution with a binomial mask).

For short boundaries a simpler curvature calculation is used. In general, the first and the last two pixels of a boundary are zero.

Grain triple junctions

Grain triple junctions are parametrized by the angles enclosed by adjacent grain boundaries. Calculating these angles requires an approximate description of the grain boundary's direction. The present implementation averages a certain number of pixels along the grain boundaries yielding tangents which are used for angle calculation.

Evaluation of parameters/results

For characterization of recrystallization dynamics different parametrizations must be taken into account. A grain boundary energy driven recrystallization regime produces a grain boundary network which differs in its microstructural features significantly from another one produced by the inhomogeneous distribution of strain energy (Kipfstuhl *et al.*, 2009).

With increasing resolution the problem arises which grain sizes to consider for the average grain size of a section. Small grains either disappear or indicate formation of new grains. The minimum grain size derived from images of thin sections acquired between crossed polarizers is limited by the thickness of the section, generally between 300 and 600 μm. Grains smaller than 300–600 μm in diameter can not be detected. Considering a certain number of largest grains has been the only possibility to compare different sections (e.g. Gow, 1969). In the present case, the quality of grain boundaries in the images is far too high to exclude arbitrarily a fraction of small grains. The size and number of grains (above the grain size threshold) is known. Different parametrizations are possible, amongst others the most frequent grain size, the average of all grains, of certain percentages of largest grains and for reasons of comparability also of a fixed number of largest grains.

Grain size distributions (grains larger than 13 μm) have interactively been extracted for sections from 102 to 279 m depth (Fig. 13). To make comparisons easier, the distribution normalized by mean grain size is also shown. Both the relative occurrence of grains larger than 10 mm and of grains smaller than 60 μm increase from 102 m to 279 m. This complex

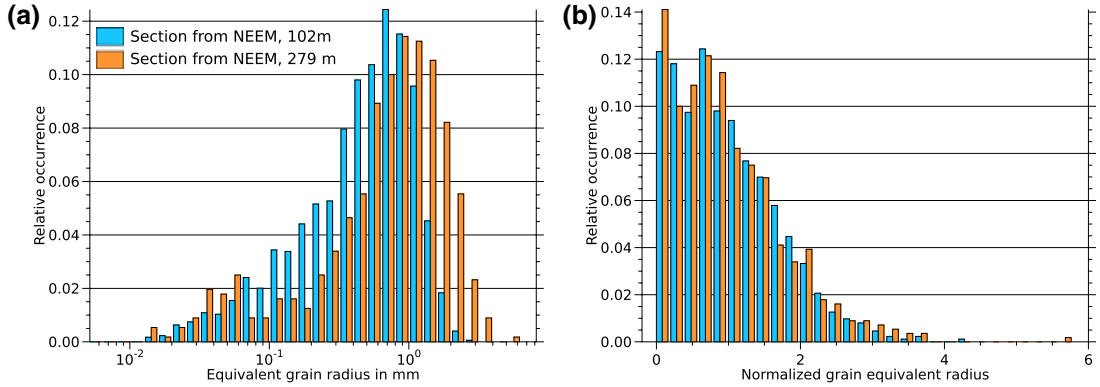

Fig. 13. Two grain size distributions at 102 and 279 m with grains larger than 20 pixels. Distribution of equivalent grain radius unnormalized (a), normalized by mean grain size (b).

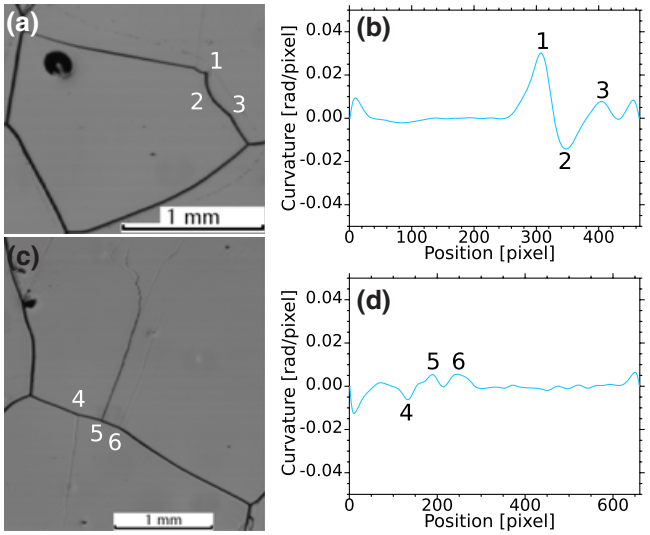

Fig. 14. Local curvature of grain boundaries. Local maxima of curvature are frequently correlated with the occurrence of subgrain boundaries (4, 5, 6) (NEEM, 322 m).

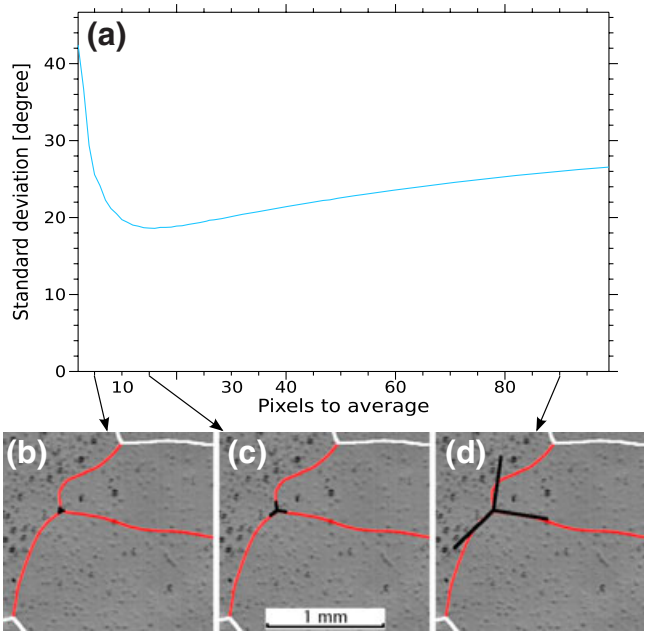

Fig. 15. (a) Mean standard deviation of planar angles versus number of pixels used to determine the tangents of two adjacent grain boundaries (2700 junctions from 322 m depth), (b)–(d) dependence on 'averaging range' for one junction (5, 15 and 90 pixels).

process cannot be described by one single parameter mentioned before.

The curvature of a grain boundary provides an estimate of the lower bound of the stored strain energy and the dislocation density which is hard to measure in large ice samples (De La Chapelle *et al.*, 1998). Local curvatures of grain boundaries are plotted in Figure 14.

The distribution of planar angles may help to quantify the transition from grain boundary energy driven grain boundary migration to one induced by strain energy. Broadening of planar angles in firn with increasing depth has previously been observed (e.g. by Kipfstuhl *et al.*, 2009).

The extracted angles depend significantly on the number of pixels considered for averaging if grain boundaries have irregular shape (Fig. 15). At low numbers (less than 10 pixels), the discretization error is large due to a low number of possible boundary shapes, which causes (in average) higher standard deviation than for values between 10 and 20. The more pixels are considered the more irregular grain boundaries appear which increases the mean standard deviation. The change in standard deviation with depth is much smaller than between different ranges used for averaging as shown in Figure 16. The shape of the curve in Figure 15 is similar for different depths, however, the absolute values are different.

Instead of choosing an arbitrary number of pixels, the approach of Chandross & Holm (2010) can be adopted to find the best approximation of each grain boundary based on Pearson's correlation coefficient.

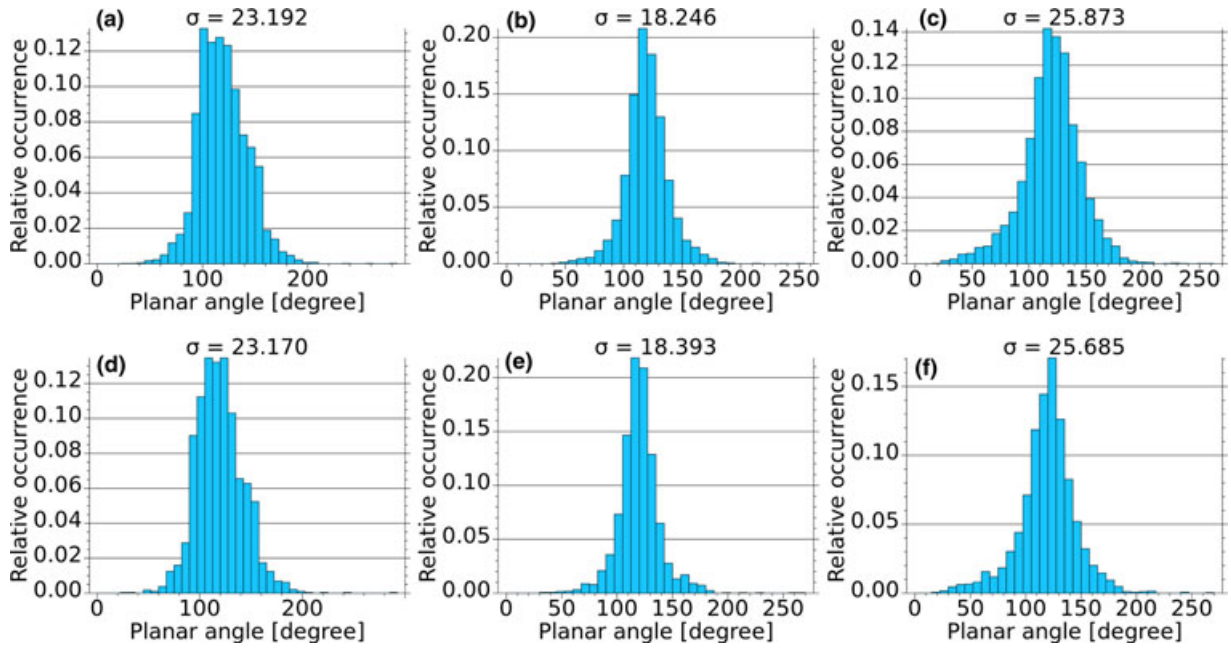

Fig. 16. Distribution of angles in grain triple junctions for sections from 102 m (a)–(c) and 279 m (d)–(f) depth (same as used in Fig. 13), different 'averaging ranges' applied: 5 pixels (a)/(d), 15 pixels (b)/(e), 90 pixels (c)/(f).

Conclusions

A dedicated method of automatic image analysis for extraction and parametrization of grain boundary networks, as they appear as result of controlled sublimation on the surface of polished ice samples, has been developed. It allows for automatic processing of large data sets as available from deep polar ice cores. Grain boundary networks can be extracted in high resolution from microstructure images. For the first time, small grain sizes and local curvatures of grain boundaries can systematically be investigated. The obtained parametrization contributes to a better understanding of the deformation and recrystallization processes on the grain scale. The image analysis method may find future applicability for other microstructure image types, for example, using thermal etching (Mullins, 1957), in which information on crystal orientation cannot be utilized for extraction of grains, but instead grain boundaries are mapped directly.

Acknowledgements

We gratefully acknowledge support from the German Science Foundation (DFG) through the grant GA1271/8-1 within the Priority Program SPP-1158. The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the manuscript. NEEM is directed and organized by the Center of Ice and Climate at the Niels Bohr Institute and US NSF, Office of Polar Programs. It is supported by funding agencies and institutions in Belgium (FNRS-CFB and FWO), Canada (NRCan/GSC), China (CAS), Denmark (FIST), France (IPEV, CNRS/INSU, CEA and ANR), Germany (AWI), Iceland (RannIs), Japan (NIPR), Korea (KOPRI), The Netherlands (NWO/ALW), Sweden (VR), Switzerland (SNF), United Kingdom (NERC) and the USA (US NSF, Office of Polar Programs).

References

Alley, R. (1992) Flow-law hypotheses for ice-sheet modeling. *J. Glaciol.* **38**, 245–256.

Arnaud, L., Gay, M., Barnola, J.M. & Duval, P. (1998) Imaging of firn and bubbly ice in coaxial reflected light: a new technique for the characterization of these porous media. *J. Glaciol.* **44**, 326–332.

Beck, P. & Sperry, R. (1950) Strain induced grain boundary migration in high purity aluminium. *J. Appl. Phys.* **21**, 150–152.

Breiman, L. (2001) Random forests. *Mach. Learn.* **45**, 5–32.

Budd, W. & Jacka, T. (1989) A review of ice rheology for ice sheet modelling. *Cold Reg. Sci. Technol.* **16**, 107–144.

Calle, M. & Urrea, V. (2010) Letter to the Editor: Stability of Random Forest importance measures. *Brief. Bioinform.* **12**, 86–89.

Chandross, M. & Holm, E. (2010) Measuring Grain Junction Angles in Discretized Microstructures. *Metall. Mater. Trans.* **41**, 3018–3025.

Cormen, T., Leiserson, C., Rivest, R. & Stein, C. (2009) *Introduction to Algorithms*. The MIT Press, 3rd edn. Cambridge, Massachusetts.

De La Chapelle, S., Castelnau, O., Lipenkov, V. & Duval, P. (1998) Dynamic recrystallization and texture development in ice as revealed by the study of deep ice cores in Antarctica and Greenland. *J. Geophys. Res.* **103**, 5091–5106.

Förstner, W. (1986) A Feature Based Corresponding Algorithm for Image Matching. *Int. Arch. Photogramm. Remote Sens.* **26**, 150–166.

Gay, M. & Weiss, J. (1999) Automatic reconstruction of polycrystalline ice microstructure from image analysis: application to the EPICA ice core at Dome Concordia, Antarctica. *J. Glaciol.* **45**, 547–554.

Gow, A. (1969) On the rates of growth of grains and crystals in south polar firn. *J. Glaciol.* **8**, 241–252.

Harris, C. & Stevens, M. (1988) A Combined Corner and Edge Detector. *Proceedings of the 4th Alvey Vision Conferenc*. Manchester, UK.

Hastie, T., Tibshirani, R. & Friedman, J. (2009) *The Elements of Statistical Learning*. Springer-Verlag, New York.

Heuer, A., Sellers, D. & Rhodes, W. (1969) Hot-working of aluminum oxide: I, primary recrystallization and texture. *J. Am. Ceram. Soc.* **52**, 468–474.

Humphreys, F. & Hatherly, M. (2004) *Recrystallization and Related Annealing Phenomena*. 2nd edn. Elsevier, Oxford, UK.

Kipfstuhl, S., Hamann, I., Lambrecht, A., Freitag, J., Faria, S., Grigoriev, D. & Azuma, N. (2006) Microstructure mapping: a new method for imaging deformation-induced microstructural features of ice on the grain scale. *J. Glaciol.* **52**, 398–406.

Kipfstuhl, S. *et al.* (2009) Evidence of dynamic recrystallization in polar firn. *J. Geophys. Res.* **114**, B05204.

Köthe, U. (1995) Primary image segmentation. *Mustererkennung, 17th DAGM Symposiu.* (ed. by G. Sagerer, S. Posch and F. Kummert), pp. 554–561. Springer-Verlag, London.

Köthe, U. (2003a) Edge and junction detection with an improved structure tensor. *Pattern Recognition, 25th DAGM Symposium, Springer LNCS 2781* (ed. by B. Michaelis and G. Krell), pp. 25–32. Magdeburg, Germany.

Köthe, U. (2003b) Integrated Edge and Junction Detection with the Boundary Tensor. *Proceedings of the 9th International Conference Computer Vision*, pp. 424–431. Nice, France.

Lindeberg, T. (1994) *Scale-Space Theory in Computer Vision*. Kluwer. Dordrecht, Netherlands.

Mueller, H. (1935) Nature of Recrystallization: I–III. *Z. Physi.* **96**, 279–327.

Mullins, W. (1957) Theory of thermal grooving. *J. Appl. Phys.* **28**, 333–339.

Ren, X. & Malik, J. (2003) Learning a classification model for segmentation. *Proceedings of the 9th International Conference Computer Visio*. pp. 10–17. Nice, France.

Tschumperlé, D. (2006) Fast anisotropic smoothing of multi-valued images using curvature-preserving PDE's. *Int. J. Comput. Visio.* **68**, 65–82.

Ueltzhöffer, K., Bendel, V., Freitag, J., Kipfstuhl, S., Wagenbach, D., Faria, S. & Garbe, C. (2010) Distribution of air bubbles in the EDML and EDC (Antarctica) ice cores, using a new method of automatic image analysis. *J. Glaciol.* **56**, 339–348.

Vincent, L. & Soille, P. (1991) Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 583–598.

Vischer, N., Huls, P. & Woldringh, C. (1994) Object-image: an interactive image analysis program using structured point collection. *Binar.* **6**, 160–166.

Weikusat, I., Miyamoto, A., Faria, S., Kipfstuhl, S., Azuma, N. & Hondoh, T. (2011) Subgrain boundaries in Antarctic ice quantified by X-ray Laue diffraction. *J. Glaciol.* **57**, 111–120.