Investigation of a Hybrid Algorithm for Sea Ice Drift Measurements Using Synthetic Aperture Radar Images

Anders Berg and Leif E. B. Eriksson, *Member, IEEE*

Abstract—Areal matching by phase correlation and feature tracking are two complementary methods used to measure sea ice drift between synthetic aperture radar images. This paper evaluates a new algorithm that combines the two methods. Areal matching is improved by new methods to handle large motions and rotated ice. It is shown that areal rotation can be resolved using a frequency-domain approach. Image segmentation is a prerequisite for feature tracking and achieved by a new method that performs better than Otsu's method for two-component Gaussian mixture distributions. A circular weighted median filter is found to be suitable for the filtering of the motion field. The algorithm is evaluated through a thorough analysis of the response and sensitivity to various algorithm settings. The accuracy of the algorithm varies by up to 50% for one image pair within the studied range of parameter settings, thus indicating the need for a proper initialization of the algorithm.

Index Terms—C-band, feature tracking, phase correlation, rotation, sea ice motion, segmentation, synthetic aperture radar (SAR).

I. INTRODUCTION

FROM one day to the next, an arbitrary ice floe in the Fram Strait may drift beyond 50 km in any direction. The same floe must thus be tracked within a search area of nearly $8000 \, \text{km}^2$. This infers a great challenge for any motion tracking algorithm that also has to be stable and highly efficient.

A few methods have been developed for the purpose of mapping sea ice drift from satellite imagery. The first algorithms appeared in the late 1980s and were applied to passive microwave data from instruments such as the Special Sensor Microwave/Imager and synthetic aperture radar (SAR) data from, e.g., Seasat. Several algorithms were developed based on the maximum cross-correlation (MCC) method [1]–[4]. Another area-based technique is known as the optical flow method and has been demonstrated in [5].

The MCC method was taken one step further with the introduction of phase correlation [6] in sea ice tracking [7], [8]. Phase correlation resembles cross-correlation but is faster as it makes benefit of the efficient fast Fourier transform via the circular convolution theorem. While in the spatial frequency domain, the phase information is preserved, and the magnitude is normalized in order to achieve illumination invariance. In [8], a hierarchical multiresolution processing system is utilized. Downsampled images and large target windows are used in the initial stage of the algorithm to find a draft motion field. The motion field is used to align target windows to cover the same ice even though the geographic location may differ. It is then iteratively refined using higher resolution and smaller windows. A basic assumption in this algorithm is that one dominant translation, or global motion, can be estimated between the two images. For images covering large areas or highly dynamical ice fields, this estimate will not always give a valid representation of the true motion. Furthermore, the study does not examine the effect on the accuracy from downsampling the images. The motion tracking approach established in [8] has been adopted in, e.g., [9]–[11], in which it has been evaluated in the Weddell Sea as well as the Baltic Sea.

In contrast to area-based techniques stand feature tracking approaches. Tracking the outlines of sea ice floes can be made using, for example, a probability model that determines how features in one image will appear in a latter image [12]. Another statistical approach takes into account certain characteristics such as the floe area, geometry, and location with reference to neighboring floes and performs a constrained search for matching ice floes [13].

In [14], the MCC method is used in a hybrid algorithm together with a feature tracking approach. The cross-correlation technique is intended to be applied within the pack ice where the motion is mostly translational with small rotation. The rotation is retrieved by an exhaustive search, in which one of the images is rotated in steps and the cross-correlation is computed for each rotation angle. The feature tracking approach is used within the marginal ice zone (MIZ), defined as the area of pack ice where the influence of the open ocean is directly observed [15]. The image is segmented, and the floe-lead boundaries are represented by so-called $\psi-s$ curves [16], [17], which are correlated to determine the translation and rotation for each floe. This hybrid algorithm has been applied to SAR data also in [18].

In this paper, we will loosely follow the areal tracking technique developed in [8], particularly in the sense of a hierarchical processing system, and combine it with a feature tracking approach based on the methodology described in [13]. The former method is improved by a new rotation resolving module.
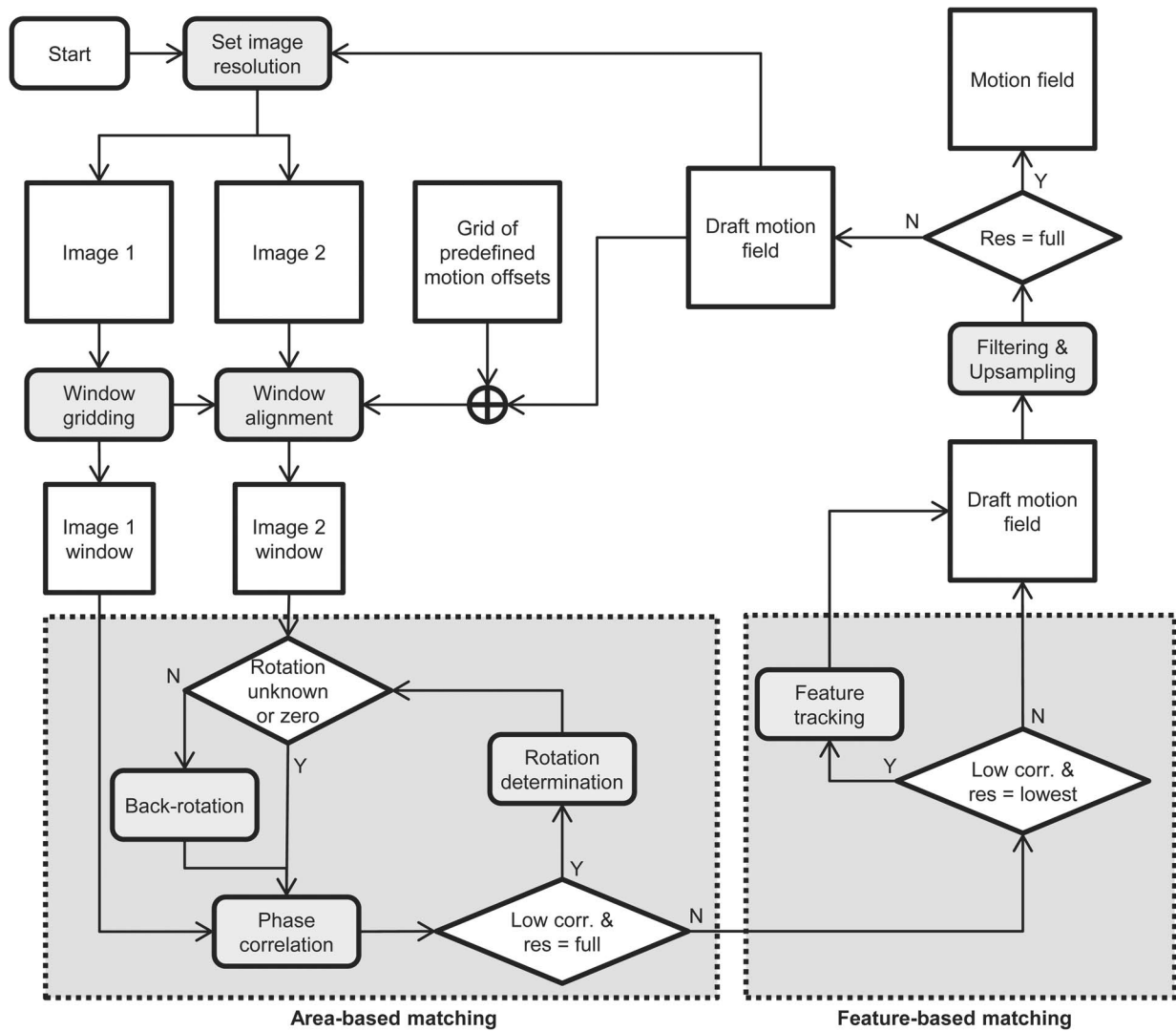
Fig. 1. Algorithm flowchart. Processes are shown in rounded rectangles, and products are shown in squares. The motion field in the upper right corner is the final product.

The latter method is preceded by a new straightforward method for image segmentation. By combining these two methods in a hybrid algorithm, the aim is to obtain a robust algorithm that is able to cope with pack ice as well as marginal ice. The algorithm is evaluated through a thorough analysis of the response and sensitivity to various algorithm settings.

The next section describes the data used in the study. Sections III and IV describe the theory, which is composed of area-based matching and feature-based tracking. The theory is followed by a section on algorithm validation and discussion. Summary and conclusions are presented in the last section.

II. DATA DESCRIPTION

The data used in this study are wide swath SAR data from the ENVISAT satellite. ENVISAT carried a C-band Advanced Synthetic Aperture Radar (ASAR) sensor, and the used images were acquired in wide swath mode. The images cover a swath of 400 km and have a spatial resolution of approximately 150 m in ground range by 150 m in azimuth and a pixel spacing of 75 m by 75 m. The images have been acquired over the

Fram Strait in 2011 and 2012. All images were acquired with horizontal polarization for transmission and reception (i.e., HH-polarization), which is preferred for operational sea ice mapping [19].

III. AREA-BASED MATCHING

A. Overview of Implementation

This section aims to give an overview of the implementation of the phase-correlation method and to highlight some of the differences from the algorithm described in [8]. Fig. 1 shows how the areal-based matching is incorporated into the algorithm and gives an overview of the processing chain. Starting in the upper left corner, the images are first downsampled to the resolution used at the initial processing level. Next, the positioning of the image windows from the first image is established. This gridding is explained in more detail hereinafter. The image windows of the second image are aligned to match the image windows from the first image, using the best available information about the drift. At the initial state, no such information

is available. Instead, the image windows are colocated, and a grid of predefined offsets is used to widen the tracking range. In the next phase, the area-based matching takes place, and under the conditions of low correlation at the final processing level, a search for rotation also takes place. This is followed by the feature-based matching which is activated at the initial processing level if areal-based matching showed low correlation. The result produced by these two tracking modules is a draft motion field, which is filtered and upsampled to match the window grid used by the next processing level. The process is then iterated with a new set of parameters, including finer image resolution. When the final processing level is reached, the draft motion field will become the final product.

The rectangular grid used for positioning the image windows is initially constructed using image coordinates for the first image. The coordinates refer to the centra of the image windows that will be processed. They are converted to geographical latitude and longitude and, subsequently, to image coordinates in the second image. A check is made if some part of the window in either images appears outside the image borders or if the land fraction is higher than a threshold value. The exclusion of land areas does not only save processing time but also improve the filtering of the motion field in inshore waters. The land fraction is measured by creating a land mask based on shoreline data from the Global Self-consistent, Hierarchical, High-resolution Geography Database [20]. If the land fraction is too high, processing of the particular image windows is discontinued, and the algorithm proceeds with the next pair of windows. If not, the drift between the two image windows will be estimated using a frequency-domain approach to correlation. Once the whole grid has been processed, the resulting motion field is filtered using a circular weighted median filter [21]. The operation of the filter at position x_0 is given by

$$d_{filt}(\boldsymbol{x}_0) = \underset{d(\boldsymbol{x})}{\arg\min} \sum_{\boldsymbol{x}_i : |\boldsymbol{x} - \boldsymbol{x}_0| < R} w(\boldsymbol{x}_i) \, |d(\boldsymbol{x}_i) - d(\boldsymbol{x})| \quad (1)$$

where d is the drift component along an axis, the weighting $w(\boldsymbol{x})$ consists of phase-correlation values, and R is the filter radius. The filter is based upon the assumption that trackable features are not homogeneously spread over the image, meaning that one successfully tracked point may be the only valid measurement over a certain area. While the motion field will be regularized through filtering, the filter will also preserve the displacement zones.

The resulting motion field is inherited by the next processing level using nearest neighbor interpolation onto the denser grid that will be used next. The whole process is repeated for the denser grid, now using the draft motion field to improve the alignment of the image windows. Each processing level has its own set of parameters which, in this study, includes the window size, the step size between the windows, the median filter radius R, and a downsampling factor. The number of processing levels is determined by a user-specified parameter. This study uses a default number of three processing levels.

At the initial processing level, at the top of the hierarchical processing system, the image windows need to be large enough to cover the longest possible displacements. At the same time, they should be small enough to resolve prominent spatial variations, for example, between fast ice and drift ice. These two objectives stand against each other and constitute a problem which has not been addressed in previous studies. A solution and compromise that we have adopted is to match one window from the first image with multiple windows from the second image. One can then cover a larger area without the expense of larger windows. Each window combination will yield one candidate solution for the drift, and the one with highest correlation will be selected. The windows in the second image need to have an overlap in case the drift is toward the border of one image window. The overlap in our implementation has been set to 10% of the window side. The number of windows is dependent on the assumed maximal drift in combination with the window size, but generally, the number of target windows does not exceed five along any dimension.

B. Phase Correlation

Let $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ denote two functions (or images) on the Euclidean plane $\mathbb{R}^2$ with Fourier transforms $F_1(\boldsymbol{\xi})$ and $F_2(\boldsymbol{\xi})$, and let the second function be a translated replica of the first function such that $f_2(\boldsymbol{x}) = f_1(\boldsymbol{x} - \boldsymbol{d})$. According to the shift theorem, the Fourier transform of the second function is $F_2(\boldsymbol{\xi}) = F_1(\boldsymbol{\xi})e^{-2\pi i \boldsymbol{d} \cdot \boldsymbol{\xi}}$. The normalized cross-power spectrum is thus

$$R(\boldsymbol{\xi}) = \frac{F_1^*(\boldsymbol{\xi})F_2(\boldsymbol{\xi})}{|F_1(\boldsymbol{\xi})F_2(\boldsymbol{\xi})|} = e^{-2\pi i \boldsymbol{d} \cdot \boldsymbol{\xi}}. \quad (2)$$

The inverse Fourier transform of the right-hand side is a Dirac delta function centered at the displacement vector $\boldsymbol{d}$

$$r(\boldsymbol{x}) = \mathcal{F}^{-1}\{R(\boldsymbol{\xi})\} = \delta(\boldsymbol{x} - \boldsymbol{d}). \quad (3)$$

By finding the argument that maximizes $r(\boldsymbol{x})$, one will obtain the displacement $\boldsymbol{d}$. It shall be noted that $r(\boldsymbol{x})$ may contain several peaks in the same order of magnitude, for example, if there is an area of ice floes drifting at different speeds. In every case, the largest peak will be selected as it likely represents one of the dominating candidate solutions. Another issue with impact on the phase correlation is speckle noise, which gives rise to noise also in $r(\boldsymbol{x})$. To reduce this noise, a Gaussian low-pass filter is applied to $r(\boldsymbol{x})$. The filtering is beneficial also in terms of peak detection. One peak is, in general, spread over several pixels, and the filter will aggregate the contributions from many pixels so that the peak is enhanced.

C. Rotation Determination

In order to determine the rotation between two images, we utilize the Fourier–Mellin transform under which images exhibit translation, rotation, and scaling invariant properties [22]. Suppose that function $f_2(\boldsymbol{x}) = f_2(x, y)$ is a translated and rotated replica of function $f_1(x, y)$. Then

$$f_2(x, y) = f_1(x \cos \theta_0 + y \sin \theta_0 - d_x,$$
$$-x \sin \theta_0 + y \cos \theta_0 - d_y) \quad (4)$$

where the rotation θ_0 is measured counterclockwise. The Fourier transform of $f_2(x, y)$ is

$$F_2(u, v) = e^{-2\pi i(d_x u + d_y v)}$$
$$\cdot F_1(u \cos \theta_0 + v \sin \theta_0, -u \sin \theta_0 + v \cos \theta_0). \quad (5)$$

The effects of translation and rotation have been separated into a phase shift and a rotation of the spectrum by the angle θ_0. The log-polar coordinate transformation given by

$$\begin{cases} \rho = \log \sqrt{u^2 + v^2} \\ \theta = \arctan v/u \end{cases}$$

yields the following expression for the magnitude of the Fourier transforms:

$$|F_2(u, v)| = |F_2'(\rho, \theta)| = |F_1'(\rho, \theta - \theta_0)|. \quad (6)$$

Note that the rotation is converted to a translation by the log-polar transformation. The rotation can then be determined using, for example, phase correlation. In this paper, the assumed approach will be to use the rotation signature defined by [22]

$$\phi_1(\theta) = \int J\left(|F_1'(\rho, \theta)|\right) d\rho \quad (7)$$

where the subscript refers to the first image and the definition is equivalent for the second image. J is here a pointwise weighting function that weights different frequencies in order to emphasize signals related to edges in the images. The next step is to cross-correlate the two rotation signatures in order to retrieve the rotation at the angle of maximum correlation. Regard that the Fourier spectrum of real valued images exhibits a $180°$ symmetry such that $F(u, v) = F^*(-u, -v)$. As an effect, there will be a $180°$ ambiguity in the retrieved rotation. The ambiguity can be resolved by performing phase correlation two more times—the second image is rotated by each possible rotation angle, using an affine map, and the rotation that gives the highest correlation is selected as the most probable one.

In the finite case, (7) is a sum over discrete points from the radius of 1 pixel to half of the side length of the image window being studied. The sum is then taken over the largest circle enclosed by the image window. The choice of side length of the image window is critical. A large side length gives higher angular resolution and more data to form a more unique signal that is suitable for correlation measurements. On the other hand, it will increase the risk for an ambiguous signal. If, for example, one fraction of the ice is rotating and the other fraction is motionless, the algorithm will face two competing solutions and must select between them. This suggests that the window size should be as small as possible without losing diversity and increasing the risk of false matches.

The rotation can only be estimated when there is an accurate estimate of the drift; otherwise, the two image windows will cover different parts of the ice field and cannot be compared. The outcome of the method is thus dependent on the existing drift estimate derived from initial processing levels. In many cases, the drift is known from the surrounding areas, and a good estimate of the drift is achieved through interpolation onto the area subject to rotation. The method is not as useful
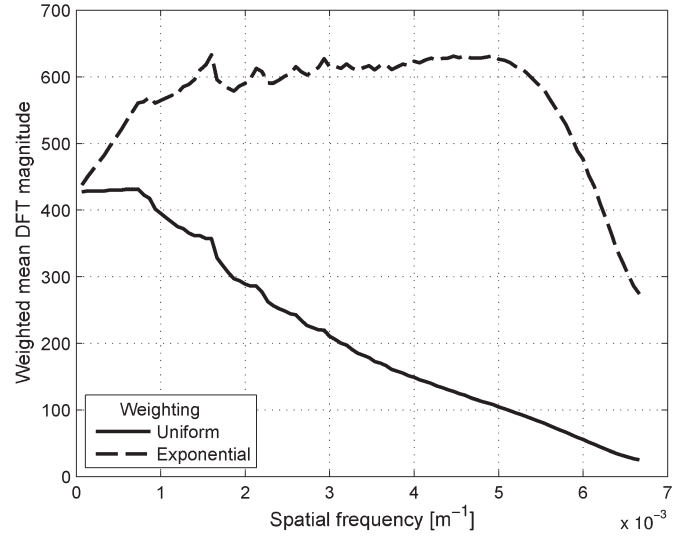

Fig. 2.   Example demonstrating the effect of applying an exponential weighting function J on the discrete Fourier transform (DFT) magnitude. The DFT magnitude here represents an average, computed from many image windows over the sea ice in an ASAR image from August 2, 2011.

in areas with lower ice concentration because the drift field is more varying and the larger fraction of open water degrades the tracking performance. It is then more likely that the image windows cover different parts of the ice field. The advantage of resolving the rotation is, however, that the local ice drift will be established, nearby regions will benefit from the drift information, and the stability of the algorithm is enhanced.

The choice of weighting function J is crucial for the performance of the rotation resolving method. The main purpose of this function is to give emphasis to Fourier–Mellin coefficients that correspond to image texture edges. The spectrum will preferably be suppressed at low frequencies where large-scale intensity shifts are represented, and the effect from discretization is notable in a log-polar coordinate system. The dc offset will be excluded so that the absolute level of the intensity is irrelevant. At high frequencies, it is desirable to suppress speckle. Based upon the mean frequency spectrum from many image windows, we adopt an exponential function of the form

$$J\left(|F_1'(\rho, \theta)|\right) = |F_1'(\rho, \theta)| \cdot e^{b\rho} \quad (8)$$

where b is a value on the order of $5 \cdot 10^{-5}$ $[1/\log(\text{m}^{-1})]$. In this case, the weighting function is to be applied to images calibrated to sigma nought and processed in decibel scale. The effect of the weighting function is demonstrated in Fig. 2. For uniform weighting, the figure shows how the energy is distributed over spatial frequencies on an average. Note that the number of samples increases with frequency, which partially explains the falloff toward higher frequency. The result was computed from all image windows over sea ice in a C-band SAR image over the Fram Strait acquired by ENVISAT on August 2, 2011.

IV. Feature-Based Matching

Feature-based matching is activated after area-based matching has been conducted. The module will be triggered if the

computed phase correlation turns out to be below a threshold value at the initial processing level. The main reason to the module only being activated at the initial processing level is that the method is suited for large image windows that are likely to provide coverage over more than a couple of ice floes. If the drift can be retrieved by feature tracking, it will replace the drift estimate based on areal matching; if not, the result from areal-based tracking will be used. The correlation threshold is a fix number which is set experimentally. The correlation over open water (where the radar signature is noiselike) is used as a reference. Naturally, all open water areas should have a correlation less than the threshold.

The method is implemented as a two-step procedure. First, the image is segmented in order to obtain a representation of the features that are present in the image. The next step is to track these features from one image to the next.

A. Image Segmentation

The images are first segmented by intensity thresholding which converts them to binary images. The resulting objects (floes) are processed to separate amalgamated floes. A set of properties, e.g., area and centroid, is determined for all the large floes. The intensity threshold is determined using a new approach developed from Otsu's threshold selection method [23]. Otsu's method selects the threshold that minimizes the weighted within-class variance. The performance of this method is unsatisfying for a bimodal distribution that is a mixture of two normal distributions with different variances. As the difference in variance increases, the method tends to select a threshold closer to the center of the mode with higher variance. To overcome this, the within-class variance is scaled with respect to the intensity range of the class. Adopting the nomenclature in [23], the threshold k is determined using the following measure of class separability:

$$\zeta(k) = \frac{\sigma_0^2}{(k-1)^2} + \frac{\sigma_1^2}{(L-k)^2}, \quad 1 < k < L. \tag{9}$$

Here, σ_0^2 and σ_1^2 are the class variances, L is the number of gray levels, and k is the threshold level. $\zeta(k)$ is a measure of the overall within-class variance with normalization to the intensity range of the respective class. For bimodal distributions, $\zeta(k)$ will typically have two distinct peaks separated by a local or global minimum at the sought threshold value. In order to avoid the selection of possible minima at the edges of the distribution, $\zeta(k)$ will not be minimized, but rather the sum of $\zeta(k)$ and the convex envelope of $-\zeta(k)$. The convex envelope is the largest possible convex underestimator of $-\zeta(k)$ over the set of intensities $K = \{1, \ldots, L\}$ and is defined as

$$\eta(k) = \sup \{c(k) : c(k') \le -\zeta(k') \, \forall k' \in K, c \text{ is convex}\}. \tag{10}$$

The threshold is then selected as

$$k_{\text{thresh}} = \arg\min_k \left(\zeta(k) + \eta(k)\right). \tag{11}$$

Fig. 3 demonstrates the application of the thresholding method along with Otsu's method on a summer SAR image
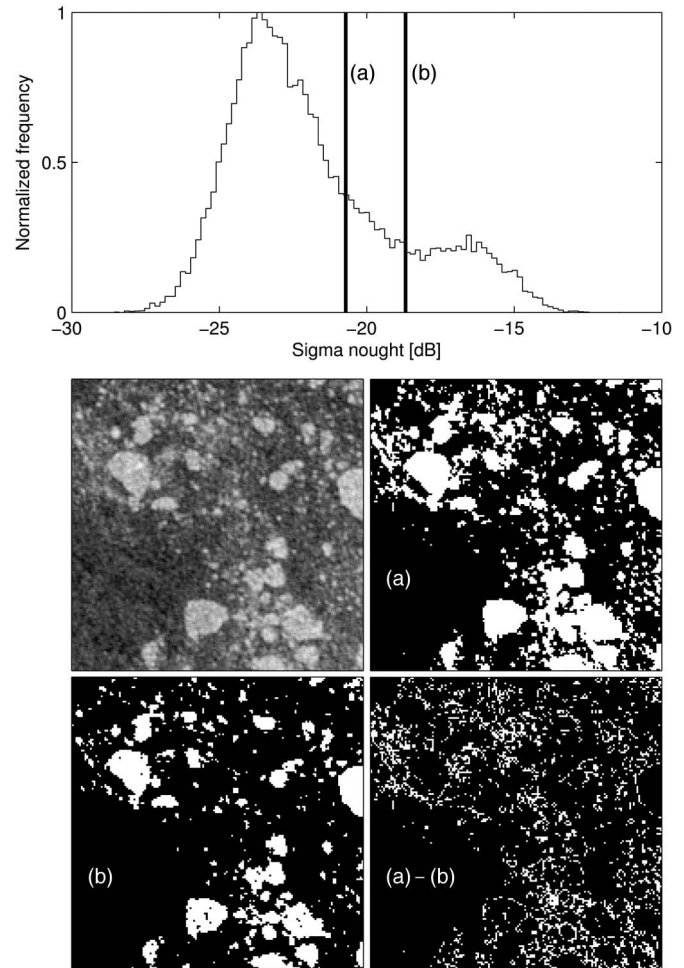

Fig. 3. Example of thresholding an image window using (a) Otsu's thresholding method and (b) the new thresholding method described in this paper. The upper panel shows the backscatter histogram with the intensity thresholds. The center left panel shows the SAR image window being analyzed, and the remaining panels show the binary image obtained for each threshold and the difference between the two binary images.

acquired by the ENVISAT satellite over the Fram Strait. While ice floes make up a fraction of the analyzed image, the largest part is open water. It can be seen that Otsu's method selects a threshold that is on the slope of the peak of the distribution that corresponds to open water pixels, whereas the new threshold cuts the distribution close to the local minimum. The binary image obtained by the thresholding will be further processed before tracking is initiated. The subsequent step is to remove all the small floes that are too difficult and not meaningful to track. That is simply done by removing all floes with an area less than a threshold of, e.g., 5 km². Smaller floes that are connected to larger floes will be removed by first making a binary erosion using linear structuring elements in horizontal, vertical, and diagonal directions. The binary erosion will disengage the small floes from the large ones and enable their removal. The contours of all large floes are restored through a binary dilation once the small floes have been removed. A morphological closing operation is performed in order to fill in small gaps within floes. After this, all floes that are connected to the image edges are also removed because their boundaries are unknown.
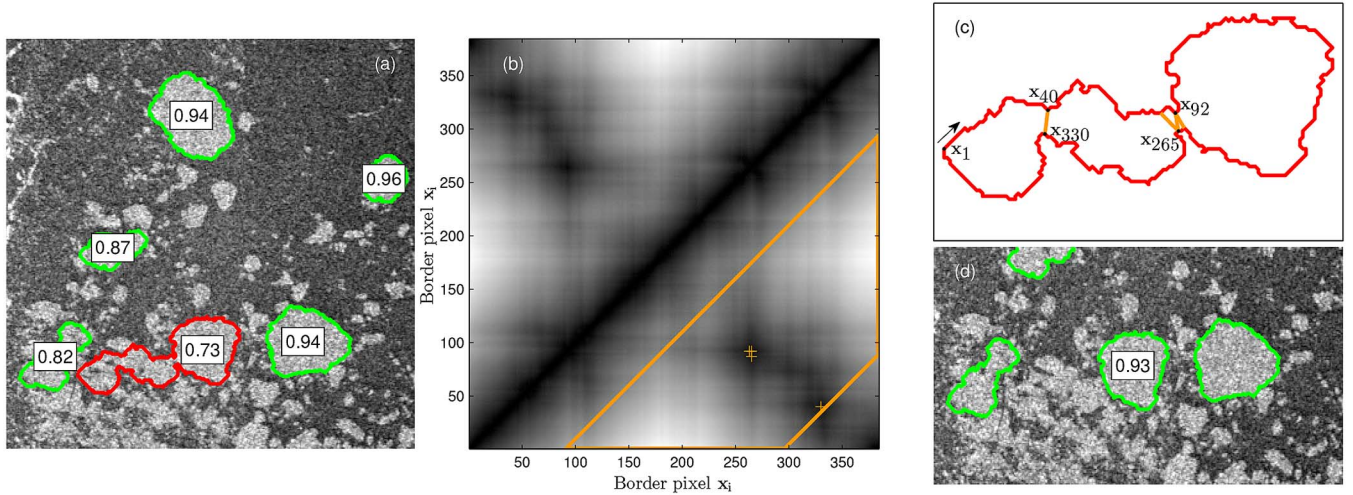
Fig. 4. Example showing the process of splitting amalgamated floes. (a) Identified floes and their respective solidities. One floe has low solidity and is selected as a candidate for splitting. (b) Distance matrix representing distances between all boundary pixels (denoted by $\mathbf{x}_i$) of the red floe, with plus signs marking local minima. Each indicated minimum corresponds to a cut between two border pixels, and the trapezoid defines a region of valid cuts. (c) Red floe with cuts. (d) Final result after the red floe was split and small floes were neglected.

One more attempt will be made to separate amalgamated floes in a process illustrated by Fig. 4. The amalgamated floes are first identified by their low solidity, which is defined as the ratio between the area of the floe and the area of its convex hull. For a floe with low solidity, the distances between all pixels on the boundary are computed, and a search is made for local minima whose distances are shorter than a threshold value. A minimum will occur at the waist of two merged floes where they are just loosely holding together. Two criteria are then used to determine whether the floe will be cut into two: The distance from edge to edge must be below a threshold value, and the perimeters of the two emerging floes must be longer than a minimum length.

B. Feature Tracking

The approach used for feature tracking is, in all essentials, based on [13] and will only be described briefly. The coordinates that outline a specific floe at time t_0 are denoted by $\mathbf{x}_i$, $i = 1, 2, \ldots, m$, and those at time t_1 are denoted by $\mathbf{y}_i$, $i = 1, 2, \ldots, n$. The drift is estimated floe by floe by iteratively solving a pair of minimization problems. First, a subset of the coordinates $\mathbf{x}_i$ is determined so that each coordinate $\mathbf{y}_i$ has a matching coordinate $\mathbf{x}_{j(i)}$. More specifically, $\mathbf{x}_{j(i)}$ is selected as the $\mathbf{x}_r$ that minimizes $\|\mathbf{y}_i - \alpha - \beta \mathbf{x}_r\|_1$. Included in the norm are the translation α and the rotation matrix β, which are given some initial values based on the distance between floe centroids and an arbitrary rotation. Second, the translation and rotation parameters are estimated by minimizing $\sum_{i=1}^{n} \|\mathbf{y}_i - \alpha - \beta \mathbf{x}_{j(i)}\|_1$. The parameter estimates found in this first iteration are used to refine the selection of $\mathbf{x}_{j(i)}$ in the next iteration.

The method to deal with broken and amalgamated floes described in [13] has not been adopted because it only applies to consolidations of not more than two floes. As described in Section IV-A, an extensive treatment of amalgamated floes is instead carried out during the segmentation.

The average residual is used as a metric to evaluate a fit between two floes. It is defined as

$$AR = \frac{\sum_{i=1}^{n} \left\| \mathbf{y}_i - \hat{\alpha} - \hat{\beta} \mathbf{x}_{j(i)} \right\|}{2n - 3} \tag{12}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the estimated translation and rotation parameters.

V. ALGORITHM VALIDATION AND DISCUSSION

A. Selection of Internal Parameters—Area-Based Matching

The phase-correlation method is controlled by a number of parameters that must be determined in an appropriate way. The parameters may interact in unforeseen ways and are preferably tuned to optimize the final performance of the algorithm. The internal parameters were varied in a test using two pairs of ASAR images over the Fram Strait. One pair was acquired on April 20 and 21, 2011, separated in time by 15 h. The ice displacement was measured manually at 63 locations using image analysis software and displacements ranged from 1.5 km up to 9 km. The MIZ was narrow, and most points were positioned within the dense pack ice. The second pair was acquired on November 22 and 24, 2011, with a fairly long temporal separation of 48 h. The displacement was measured in 135 locations and varied between 4 and 51 km. Most locations were selected within the extensive MIZ of rather diverse ice motion. The accuracy presented in this section is the weighted average of the 198 drift measurements, where each image pair has equal weight on the total accuracy.

The manual measurements were then compared to the output from the algorithm. In order to preserve the accuracy of the manual measurements, the algorithm's output is adapted to match the manual measurements rather than the opposite. The magnitude and direction of the computed drift is interpolated onto the positions of the reference measurements. The interpolation is performed in the spatial domain belonging to the first

TABLE I
ACCURACY OBTAINED FROM DIFFERENT SETS OF WINDOW SIZES
(SIDE LENGTHS) IN A THREE-LAYERED PROCESSING SYSTEM.
SHADED CELLS MARK NONDEFAULT VALUES. WINDOW SIZES
ARE TYPICALLY CHOSEN TO HAVE LOW PRIME FACTORS

Accuracy [%]			Window size		
Total	Pair 1	Pair 2	[pixel]		
70.6	95.2	45.9	384	192	60
68.4	95.2	41.5	320	192	60
68.2	90.5	45.9	432	216	60
67.6	93.7	41.5	288	192	60
67.5	92.1	43.0	432	192	50
67.5	92.1	43.0	432	192	64
67.5	90.5	44.4	432	192	70
67.2	93.7	40.7	432	144	60
66.8	92.1	41.5	256	192	60
66.7	90.5	43.0	432	192	56
66.5	93.7	39.3	432	384	60
66.3	90.5	42.2	432	192	68
65.9	88.9	43.0	432	192	74
65.6	88.9	42.2	432	192	60
65.2	88.9	41.5	432	96	60
65.1	87.3	43.0	432	192	40

TABLE II
ACCURACY VERSUS DOWNSAMPLING FACTOR (DF). THE PARAMETERS
ARE SET INDIVIDUALLY FOR EACH PROCESSING LEVEL. INITIAL TO
FINAL PROCESSING LEVELS ARE PRESENTED FROM LEFT TO RIGHT

Accuracy [%]			DF		
Total	Pair 1	Pair 2	[-]		
67.9	93.7	42.2	2	2	1
65.6	88.9	42.2	2	1	1
63.9	93.7	34.1	1	1	1
63.4	92.1	34.8	4	2	1
63.1	92.1	34.1	4	1	1

image. The result is used to determine the new positions at the time of the second image. It can be noted that the algorithm tends to give either a correct drift estimate with relatively high accuracy or a drift that has little or no correspondence with the actual drift. Based on this fact, we measure the accuracy as the proportion of correct drift vectors. A drift vector will be considered as correct if the magnitude of the displacement is correct to within 2 km and the directional offset is less than $20°$. The condition is fairly loose as the aim is to measure the accuracy rather than the precision of the algorithm. It is a sensitive measure because each vector will either be counted or not, and therefore, it will likely generate some variation in the measured accuracy.

1) Window Side Length: A key set of parameters is the one that determines the side lengths of the image windows at each processing level in the phase-correlation algorithm. The image windows must be small in order to capture fine details in the ice motion field. They also need to be large enough to comprise a sufficient amount of image features which give the specific image window certain characteristics, making it meaningful to correlate with other windows. Table I shows the accuracy of the algorithm for different combinations of window side lengths at the different processing levels. Three processing levels have been used in all setups, and all parameters are constant, except for the window size and two more parameters: the step size and the radius of the weighted median filter. The step size, which determines the spatial density of the phase-correlation measurements, has been set to half the window size. The filter radius has been set equal to the window size. The window size has been given default values of [432 192 60] pixels for the three processing levels. These values were chosen because they were known to give decent results.

The table shows that the accuracy is moderately affected by the selection of window sizes. The highest accuracy is achieved with window sizes very similar to the default values, but the default values appear in the lower end of the scale. The window

size at the last processing level has no significant impact on the accuracy within the studied interval from 40 to 74 pixels, although a side length of 40 pixels gives the lowest accuracy and is especially troublesome for the April images which are dominated by dense pack ice. This low accuracy for small windows in pack ice may be related to a lower occurrence of features, such as ridges and cracks, compared to the ice in the MIZ where individual floes are frequent and likely to appear even in very small windows.

The window size at the intermediate processing level has a small impact on the accuracy as well. Looking at the initial processing level, window sizes just below the default value give the highest accuracy overall. A large size is preferred in order to catch the longest potential displacements (and to minimize the number of target windows—see Section III-A). Unnecessary large windows can, however, cause a problem in heterogeneous motion fields. Because only one value of the drift is obtained for each window, the measured displacement may not be representative for the whole window area.

One may note the large discrepancy in accuracy between the two image pairs, with the first pair having much higher accuracy. This is an effect of the MIZ being wider in the second image pair, which poses a challenge for the phase-correlation method.

2) Downsampling: In [8] is suggested a methodology to use downsampled images at the initial processing level and to increase the resolution by a factor of 2 until the full resolution is reached at the final processing level. The purpose of downsampling is to reduce computational load and mask out small motion. If there is a small spread in magnitude and direction among the candidate solutions, downsampling will efficiently merge these solutions. The drawback is that the ice floe edges, ridges, and other structures will be blurred and cause a reduction in the correlation. Table II shows the effect of varying the downsampling factor at the different processing levels while keeping all other parameters constant (the window size set to [432 192 60] pixels for the three processing levels, the step size set to half the window size, and the radius of the weighted median filter set equal to the window size). It seems that the impact of varying the downsampling at the intermediate processing level is small, whereas it is more sensitive at the initial processing level. High downsampling (by a factor of 4) or no downsampling gives some decrease in accuracy, whereas downsampling by a factor of 2 seems adequate. It may not be the best strategy to start with highly downsampled images—if there is an initial error in the drift estimation, it will propagate to lower levels in the hierarchical processing system.

TABLE III
ACCURACY VERSUS WINDOW-TO-STEP SIZE RATIO. THE PARAMETERS
ARE SET INDIVIDUALLY FOR EACH PROCESSING LEVEL. INITIAL TO
FINAL PROCESSING LEVELS ARE PRESENTED FROM LEFT TO RIGHT

Accuracy [%]			Window/		
Total	Pair 1	Pair 2	Step		
72.2	98.4	45.9	4	2	2
72.1	96.8	47.4	8	2	2
69.5	96.8	42.2	3	2	2
65.6	88.9	42.2	2	2	2
64.0	96.8	31.1	1.5	2	2

3) Step Size: The step size is set in relation to the window side length. A window-to-step size ratio of 2, for example, means that two adjacent windows overlap each other by half of their side length. Table III shows the effect of varying the step size while keeping all other parameters constant with the aforementioned values (and downsampling factors of [2 1 1]). The table indicates a trend toward higher accuracy for denser sampling. The relative improvement is higher for smaller ratios, and the benefit of very dense sampling must also be weighed against the computational cost. For ratios above 4, the benefit of even denser sampling is very small as the accuracy appears to saturate. The two cases with the most and least densely spaced windows are shown for one image pair in Fig. 6. When sparse spacing is used, regions of homogeneous drift appear more disconnected with sharp boundaries in between. If a drift estimate from one phase correlation is incorrect, it will more likely be filtered out in the case of more densely spaced windows. For sparsely spaced windows, an incorrect measurement may appear as an isolated rectangular region of distinct motion that differs from the surrounding area.

4) Filtering: The radius of the weighted median filter determines how many measurements should be weighted together in order to form a representative value of the drift in a certain position. We choose to express the filter radius in relative units of window side length instead of as an absolute distance. It is reasonable because the finest resolved motion is governed by the window size. The number of measurements within a certain radius is dependent on the window-to-step size ratio, which was set to 2 for all processing levels in this test. Note that the number of measurements increases stepwise with the radius. For example, as the radius exceeds $\sqrt{2}$ times the window side length, the number of filtered measurements increases from 21 to 25.

Fig. 5 shows the results from varying the radius of the median filter and keeping all other parameters constant. The window size was set to [432 192 60] for the three processing levels. It can be seen from the figure that the accuracy is less sensitive to the radius at the final processing level than at higher levels. Suboptimal accuracy is obtained for large radii, due to the spatial smearing of the motion field, and for small radii when the filter operates with few measurements. The optimal size of the filter is found with radii in the order of the window size. As in the previous test of the step size and observable in Fig. 6, the marginal ice is tracked with lower accuracy than the pack ice. It is also within the MIZ that the accuracy varies at most with respect to filter radius.
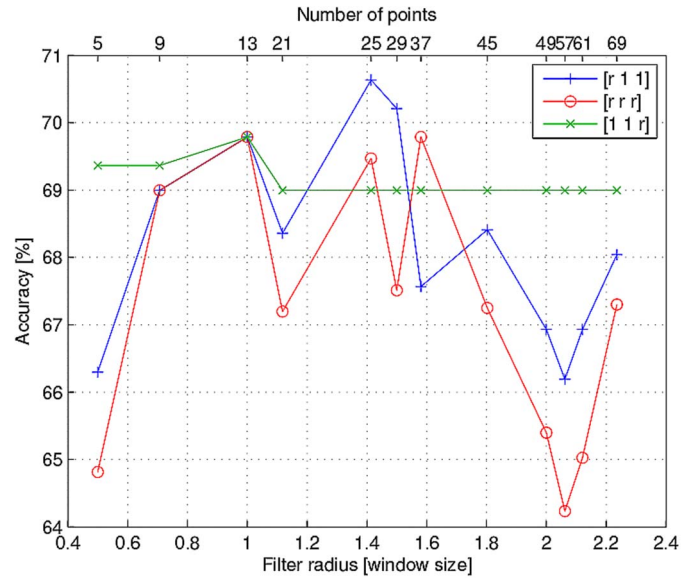

Fig. 5. Accuracy versus radius of weighted median filter. Also shown is the number of measurement points covered by the radius. Each curve corresponds to one setup, shown in the legend, where the filter radius is set to either r or 1 at the three processing levels.

5) Rotation Measurement: The performance of the rotation measurement was evaluated using the image pair acquired by ENVISAT over the Fram Strait on November 22 and 24, 2011. The rotation was measured manually in the SAR images and compared to the rotation derived by the aforementioned method. The manual measurements were carried out by measuring the position of two nearby points in one image and locating the same points in a second image. If the ice does not break between the points, the rotation can be determined as the difference in direction between the two lines that intersect the points in each image. 114 mutual points were marked out in the images and used to compute the rotation in 57 locations. The results are shown in Fig. 7. The estimated rotation is evaluated using root-median-square errors, i.e., the square root of the median squared error. It is preferred before root-mean-square error because of its robustness to outliers. The error is shown for window side lengths up to 200 pixels. The angular resolution depends on the pixel size and can be improved through the interpolation of the images prior to the Fourier–Mellin transform (at the cost of computational load). The images were oversampled by a factor of 4, which gave a slight improvement for smaller image windows, but the effect disappears when the side length goes above 60 pixels. The minimum error is reached for side lengths between 5 and 10 km (or 70 and 140 pixels). Within this range, the root-median-square error is generally $3°$–$5°$ with a minimum of $3.1°$ at a window side length of 98 pixels. There is a clear trend toward larger errors for side lengths larger than 140 pixels, presumably because the ice conditions vary too much within one image window and it is not reasonable to assume one rotation angle for the whole window. The histogram shows that a majority of the measurements were carried out on a length scale of less than 100 pixels. The distribution is related to the possibility of finding rigid ice areas of different dimensions and the frequency of such areas.
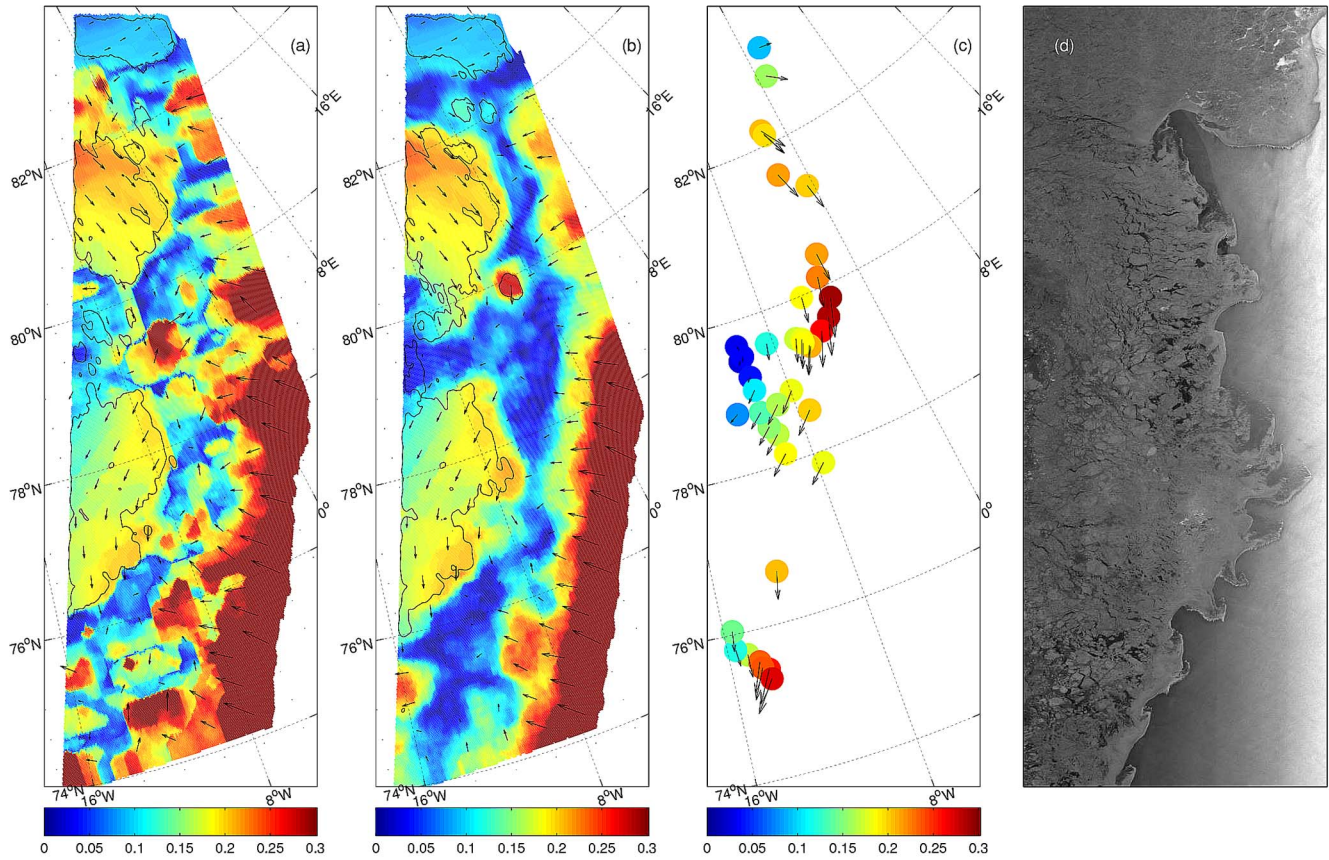
Fig. 6. Example showing the impact of varying the spacing between the image windows for which phase correlation is computed. The velocity ranges up to 0.3 m/s and is indicated by color. Contour lines mark regions of high correlation. (a) Moderately spaced windows. (b) Densely spaced windows. (c) Reference measurements, grouped by floe. (d) ASAR image from the first overpass on November 22, 2011.
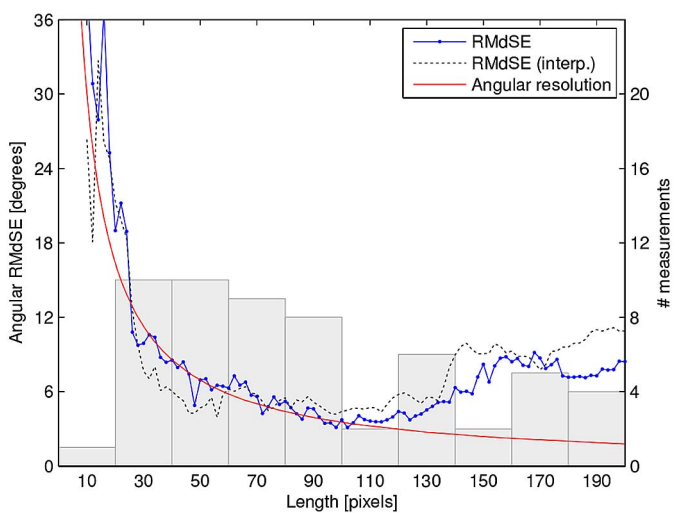

Fig. 7. Root-median-square error of the estimated rotation as a function of window side length computed from C-band SAR data with a pixel size of 75 m and for two different cases: for images in the original scale and for images oversampled by a factor of 4. The angular resolution is shown for the case without interpolation. The bar chart shows the number of rotation measurements for a certain length of rigid ice.

B. Feature-Based Matching

This section will discuss some aspects of the feature-based matching. The image segmentation is based on straightfor-

ward intensity thresholding. Other methods may, however, be considered for potential improvements of the segmentation. A number of algorithms have been developed in order to estimate the sea ice concentration from single- and dual-polarization C-band SAR data (see, for example, [24] and [25]). These algorithms may be used to distinguish sea ice from open water and perform image segmentation. It must, however, be taken into account that the ice concentration must be convertible to binary sea ice extent in the full resolution of the SAR images; otherwise, there is no possibility to outline different floes. The conversion to a binary format involves the selection of a concentration threshold; a task that may become intricate for flat ice concentration distributions. Other alternative methods to perform segmentation include using the Markov random field [26] or K-means clustering method [27]. The usage of the former method is illustrated in [28] and as a semiautomated analysis in [29]. These algorithms are able to handle the spatial relationship among pixels within the segmentation process which will reduce the sensitivity to speckle noise.

In order to determine the effect of the feature-based matching, the algorithm was tested with the corresponding module switched on and off. The test was made with four pairs of ASAR images over the Fram Strait from June 8 and 9 and August 11–16 of 2011. The computed drift was compared to 45 manual measurements, following the methodology in
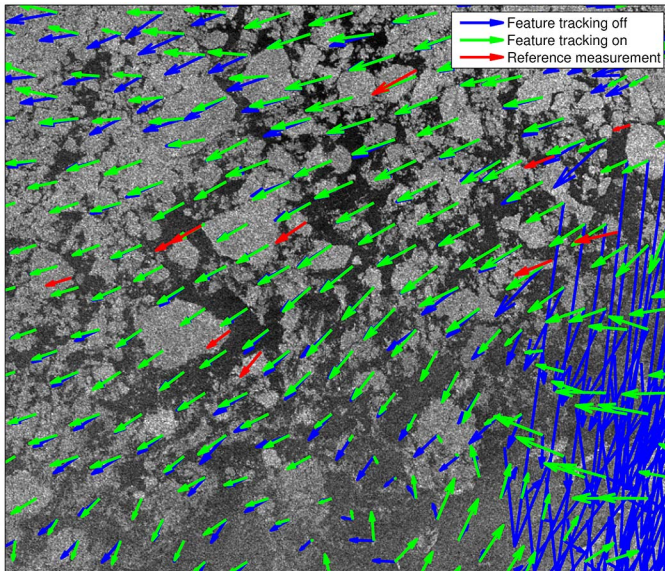
Fig. 8. Motion field computed by the algorithm with feature-based matching module switched on and off, on top of the ASAR image from August 11, 2011. Manual measurements of the ice drift are shown by red arrows. Switching off the module causes erroneous drift in the eastern part of the scene.

Section V-A. All the manual drift measurements were selected within the MIZ where the ice concentration was less than about eight-tenths. An example of the result is shown in Fig. 8. The resulting accuracy was 36% when the feature-based matching was switched off and 49% when the module was switched on. Here, it should be noted that the absolute values are dependent on the selection of reference points and the used measure of accuracy. The accuracy is defined as before, as the proportion of correct drift vectors (see Section V-A). The relative improvement created by the feature-based matching shows that the module gives valuable support to the algorithm. The algorithm works reliably with feature-based matching being subordinate to area-based matching, and the possibility for the former module to fall back on the latter assures that the level of accuracy is maintained or increased. There are still problematic regions where both the area-based and the feature-based matching fail. Typical examples are regions of heavy deformation and ice in melt conditions near the ice edge.

Run time measurements were carried out with and without feature-based matching. The run time is largely dependent on the parameter settings (primarily the step size, described in Section V-A3) and the size of the overlap between the two images. With feature-based matching switched on, the run time typically increases by 10% to 25%. On a standard personal computer and with optimally selected parameters, the run time is about 30 min for an intersecting area of 400 times 400 km^2 between two ASAR images.

VI. Summary and Conclusion

In order for an ice drift algorithm to handle the dynamics of marginal ice as well as pack ice, a hybrid algorithm is necessary. By combining areal matching with feature tracking, it is possible to determine the drift within the pack ice as well as in the MIZ. The areal matching technique used in this study is based on phase correlation. It has been extended with a module that estimates the rotation of the ice. The rotation is determined by a cross-correlation of the rotation signatures obtained from the log-polar Fourier–Mellin transform. It was shown that the rotation signatures are aptly weighted with exponential functions in order to give a balanced representation of the spatial frequency content from one image window. It is also shown that the rotation is determined with higher accuracy for a certain range of window sizes. For small windows, the angular resolution is poor, and for large windows, the rotation may become ambiguous as the motion field is more heterogeneous.

The phase correlation is computed within a hierarchical processing system which uses a number of internal parameters that set, for example, the window size at each processing level. We have investigated how some key parameters affect the accuracy of the algorithm. The sensitivity to window size is fairly low which means that other factors than the accuracy, such as the longest potential displacement, should be considered for the window size selection.

The algorithm uses downsampling to shorten the processing time and filter out small motion. It is shown that the amount of downsampling at the initial processing level has an impact on the accuracy, whereas intermediate levels are not affected to the same extent. The spacing between image windows is also shown to influence the accuracy, with higher accuracy for shorter spacing. Based on the results in this study, it appears that the effect saturates when the window-to-step size ratio is around 4. Using a higher ratio is not motivated by the relatively small improvement and, not less important, by the high computational cost. The last parameter that has been investigated is the radius of the weighted median filter. It appears that the most suitable choice is in the same order as the window size. The filter is then likely to filter out erroneous drift measurements without degrading the motion field through spatial smearing.

The influence of the feature tracking module was tested against manual drift measurements and proved to have a positive effect within the MIZ. The additional computational load is well matched by the increase in accuracy. Future work may consider other methods to perform image segmentation which may be able to handle a larger variety of ice conditions.

The perspective of this study has been centered toward the response of the algorithm to internal parameter settings. Future studies may provide more knowledge about the sensitivity to temporal spacing between images, season, weather conditions, and study region. This study will then provide an important foundation for further analyses.

REFERENCES

[1] M. Fily and D. Rothrock, "Sea ice tracking by nested correlations," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-25, no. 5, pp. 570–580, Sep. 1987.

[2] T. Hirose, T. Heacock, and F. Lee, "Advances in sea ice tracking algorithms," in *Proc. 14th Can. Symp. Remote Sens.*, 1991, pp. 292–295.

[3] W. Emery, C. Fowler, J. Hawkins, and R. Preller, "Fram Strait satellite image-derived ice motions," *J. Geophys. Res.*, vol. 96, no. C3, pp. 4751–4768, Mar. 1991.

[4] Y. Sun, "A new correlation technique for ice-motion analysis," *EARSeL Int. J. Adv. Remote Sens.*, vol. 3, no. 2, pp. 57–63, 1994.

[5] Y. Sun, "Automatic ice motion retrieval from ERS-1 SAR images using the optical flow method," *Int. J. Remote Sens.*, vol. 17, no. 11, pp. 2059–2087, Jul. 1996.

[6] F. Pla and M. Bober, *Estimating Translation/Deformation Motion Through Phase Correlation,* 1st ed. Berlin, Germany: Springer-Verlag, 1997, ser. Lecture Notes in Computer Science, ch. 7, pp. 653–660.

[7] M. Thomas, "Global motion estimation of sea ice using synthetic aperture radar imagery," M.S. thesis, Univ. Delaware, Newark, DE, USA, 2004.

[8] M. Thomas, C. A. Geiger, and C. Kambhamettu, "High resolution (400 m) motion characterization of sea ice using ERS-1 SAR imagery," *Cold Regions Sci. Technol.*, vol. 52, no. 2, pp. 207–223, Apr. 2008.

[9] T. Hollands and W. Dierking, "Performance of a multiscale correlation algorithm for the estimation of sea-ice drift from SAR images: Initial results," *Ann. Glaciol.*, vol. 52, no. 57, pp. 311–317, May 2011.

[10] A. Berg, L. Eriksson, K. Borens, and H. Lindh, "Observations and Analysis of Sea Ice Motion With the Ice Buoy DRIVA During the 2010 Spring Field Campaign in the Bay of Bothnia," Dept. Earth Space Sci., Chalmers University of Technology, Göteborg, Sweden, Tech. Rep. 7, 2011.

[11] J. Karvonen, "Operational SAR-based sea ice drift monitoring over the Baltic Sea," *Ocean Sci.*, vol. 8, no. 4, pp. 473–483, Jul. 2012.

[12] J. Banfield, "Automated tracking of ice floes: A stochastic approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 6, pp. 905–911, Nov. 1991.

[13] S. Das Peddada and R. McDevitt, "Least Average Residual Algorithm (LARA) for tracking the motion of Arctic sea ice," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 915–926, Jul. 1996.

[14] R. Kwok, J. Curlander, R. McConnell, and S. Pang, "An ice-motion tracking system at the Alaska SAR facility," *IEEE J. Ocean. Eng.*, vol. 15, no. 1, pp. 44–54, Jan. 1990.

[15] M. Leppäranta, *The Drift of Sea Ice*. Berlin, Germany: Springer-Verlag, 2011.

[16] D. Ballard and C. Brown, *Computer Vision*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1982.

[17] R. McConnell, R. Kwok, J. Curlander, W. Kober, and S. Pang, "$\Psi - S$ correlation and dynamic time warping: Two methods for tracking ice floes in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 6, pp. 1004–1012, Nov. 1991.

[18] R. Kwok, A. Schweiger, D. Rothrock, S. Pang, and C. Kottmeier, "Sea ice motion from satellite passive microwave imagery assessed with ERS SAR and buoy motions," *J. Geophys. Res.*, vol. 103, no. C4, pp. 8191–8214, Apr. 1998.

[19] W. Dierking and L. T. Pedersen, "Monitoring sea ice using Envisat ASAR—A new era starting 10 years ago," in *Proc. IEEE IGARSS*, Munich, Germany, 2012, pp. 1852–1855.

[20] P. Wessel and W. Smith, "A global, self-consistent, hierarchical, high-resolution shoreline database," *J. Geophys. Res.*, vol. 101, no. B4, pp. 8741–8743, Apr. 1996.

[21] L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo, "Weighted median filters: A tutorial," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 43, no. 3, pp. 157–192, Mar. 1996.

[22] M. McGuire, "An Image Registration Technique for Recovering Rotation, Scale and Translation Parameters," NEC Res. Inst., Tokyo, Japan, Tech. Rep. 98-018, 1998.

[23] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 285–296, Jan. 1979.

[24] A. Berg and L. E. B. Eriksson, "SAR algorithm for sea ice concentration evaluation for the Baltic Sea," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 5, pp. 938–942, Sep. 2012.

[25] J. Karvonen, "Baltic sea ice concentration estimation based on C-band HH-polarized SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 6, pp. 1874–1884, Dec. 2012.

[26] R. Kindermann and J. L. Snell, *Markov Random Fields and Their Applications*, vol. 1. Providence, RI, USA: AMS, 1980.

[27] E. W. Forgy, "Cluster analysis of multivariate data: Efficiency versus interpretability of classifications," *Biometrics*, vol. 21, no. 3, pp. 768–769, Sep. 1965.

[28] H. Deng and D. A. Clausi, "Unsupervised segmentation of synthetic aperture radar sea ice imagery using a novel Markov random field model," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 528–538, Mar. 2005.

[29] Y. Qiyao and D. A. Clausi, "SAR sea-ice image analysis based on iterative region growing using semantics," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3919–3931, Dec. 2007.

Anders Berg received the M.Sc. degree in engineering physics and the Licentiate of Technology degree in radio and space science from the Chalmers University of Technology, Gothenburg, Sweden, in 2008 and 2011, respectively. He is currently working as a doctoral student in the Radar Remote Sensing group at Chalmers' Department of Earth and Space Sciences.

His main research topic is sea ice monitoring using spaceborne synthetic aperture radar.

Leif E. B. Eriksson (M'05) received a University Certificate in Space Technology from Umeâ University, Umeâ, Sweden, in 1993, the M.S. degree in electrical engineering from the Chalmers University of Technology, Gothenburg, Sweden, in 1998, and the Ph.D. degree from Friedrich-Schiller University, Jena, Germany, in 2004.

From 1999 to 2000, he was employed by the Joint Research Centre of the European Commission at the Space Application Institute in Ispra, Italy, where he was working with European Remote Sensing satellite data for forest and wetland mapping in Siberia. Between 2000 and 2004, he was with the Department of Geoinformatics at the Friedrich-Schiller University, where he was working with C-band and L-band SAR interferometry for forestry applications. Since 2004, he has been with the Department of Earth and Space Sciences at the Chalmers University of Technology, and in 2012, he was promoted to Associate Professor in radar remote sensing. His current research interests include the use of SAR for forestry and oceanography applications.