Manuscript prepared for J. Name
with version 2015/09/17 7.94 Copernicus papers of the LaTeX class copernicus.cls.
Date: 4 March 2016

Response to the reviewers

'Sea ice drift from Sentinel-1 SAR imagery using open source feature tracking'

S. Muckenhuber, A.A. Korosov, and S. Sandven

Nansen Environmental and Remote Sensing Center (NERSC), Thormøhlensgate 47,
5006 Bergen, Norway

Correspondence to: S. Muckenhuber (stefan.muckenhuber@nersc.no)

Dear Referees,

Thank you very much for helping us improving our paper.

Please find here the **answers** to your comments and the ***changes in the manuscript***:

1 Response to anonymous referee #1

General comments

1.1

Take full account of the availability of operational processing of Sentinel-1 SAR ice drift as processed at the Danish Technical University (DTU) and distributed through the Copernicus Marine Environment Core Service (CMEMS, former MyOcean). On page 6939 line 20 you write 'This introduces a new era in SAR Earth Observation, but no sea ice drift algorithm using Sentinel-1 data has been published so far.' Although this is correct (it is not published in peer-reviewed literature), the fact that this product exists and is the current de-facto standard for sea ice drift from Sentinel-1 SAR should reflect on the discussions of this paper. In addition to mentioning this product, the authors could use the estimates from the CMEMS product as comparison to their vectors in Figure 4 (in place or in addition to the manually drawn vectors). This product and its archive are freely accessible from http://marine.copernicus.eu

The CMEMS/DTU data set has been included and the vector field provided for the image pair 'Fram Strait' has been utilised. CMEMS/DTU data have been compared to ORB vectors and the presented manual derived drift vectors, as shown in Section 3.4 and 4.5, Figure 1 and Table 1.

1.2

The justification for a feature-based approach (as opposed to the more commonly adopted pattern-matching strategies) is only quickly discussed, mainly arguing that 'Feature-tracking provide vectors, which are independent from each other, whereas pattern recognition includes the surrounding drift information'. Here 'independent' refers to the fact that neighbouring drift vectors do not re-use the same image pixels. However, pattern-matching methods can be designed to retrieve independent vectors (varying the extent of the correlation area and the spacing between vectors). Please investigate in your revised study the spatial sampling (distance to neighbour) of your (matched) features, and relate this density to the size used for matching the features. This should lead to a discussion on these aspects of 'independence of motion vectors' with your set of parameters in feature tracking.

'Independent' refers to the two facts: the keypoints are found independently of neighbors and the features from two images are matched independently from each other. Hence, the resulting vectors are independent from each other in terms of initial position, lengths and direction and even very close drift vectors may point into different directions. Section 5 is extended by a comparison between feature tracking and pattern matching with regard to independence and resolution.

1.3

The TC reader is curious to learn more about the much higher number of features detected (and matched?) with HV than HH imagery. What is the reason for this? Where are these additional features located in an image? Are HH and HV features mostly at the same location or do they complement each others? Can the authors illustrate the distribution of HV and HH features on a (zoomed) Sentinel-1 SAR image? Are your findings the same as those of Komarov and Barber (2014), or not?

The different performance can be explained by a higher variability of the HV backscatter intensity, considering a window with the same size as used for feature description (34$\times$34 pix). The HV channel appears to be more informative compared to the HH channel which is smoother. Two figures (Figure 6 and 7) have been added. Figure 6 shows the location of detected and matched keypoints on a (zoomed) Sentinel-1 SAR image. Figure 7. shows the standard deviation distribution of the HH and HV radar backscatter. The additional results and discussion including a comparison with Komarov and Barber (2014) have been added to Section 4.2 and 5.

1.4

Make a clear distinction between Section 4 'Results' and Section 5 'Discussions'.

Sections 4 and 5 have been revised, several parts have been rewritten and new text has been added. We tried to make a more clear distinction separating the results from the discussion.

Specific comments

*) Title: although it is obvious to many, including the authors, 'open source' could be specified further as 'open source [feature tracking] software'.

The title has been changed to: ***Open source feature tracking algorithm for sea ice drift retrieval from Sentinel-1 SAR imagery***

*) Abstract: Please delete the sentence 'A new quality measure for feature tracking algorithms is introduced utilizing the resulting distribution of the vector field'. The proposed quality measure is not specific to feature tracking, and is not fully convincing (it might help detecting obviously wrong vectors, but is limited by true discontinuities of the motion fields). It is relevant to use it in your study, but not to highlight it as a new finding.

The sentence has been deleted.

*) Page 6938 line 20: could you add a sentence discussing the role of sea ice drift in the Southern Hemisphere?

We added: ***Antarctic sea ice is even more mobile and its strong seasonality is linked to the ice transport from high to low latitudes.***

*) Same page line 23-24: are you known with Komarov and Barber (2014)? The resolution they document, I believe, is 3 to 5 km. And RGPS (Kwok et al. 1990) has 5km spatial resolution. So you should probably be more specific about 'few kms' (and cite Komarov and Barber 2014). Komarov, A.S.; Barber, D.G., 'Sea Ice Motion Tracking From Sequential Dual-Polarization RADARSAT-2 Images,' in Geoscience and Remote Sensing, IEEE Transactions on , vol.52, no.1, pp.121-136, Jan. 2014 doi: 10.1109/TGRS.2012.2236845

We changed the sentence to: ***However, there is still a lack of large sea ice drift data sets with sufficient resolution to estimate convergence and divergence on a spatial scaling of less than 5 kilometres.*** **The work of Komarov and Barber (2014) has been considered in the discussion.**

*) Same page line 25: Use 'seas' instead of 'oceans'?

Agree, has been changed.

*) Citation Kwok 2010 and Kwok and Sulsky 2010: these are reviews. They 'demonstrated' these capabilities earlier (e.g. in Kwok 1990). Same for Kwok 1998 few lines below. This is where Komarov and Barber (2014) is missing.

Kwok (2010) and Kwok and Sulsky (2010) have been replaced by Kwok (1990). A sentence about Komarov and Barber (2014) has been added to the introduction: ***Komarov and Barber (2014) used similar pattern matching techniques to evaluate ice motion results from dual-polarisation Radarsat-2 images.***

*) Page 6939 line 18: although it is not published in peer-review literature, the operational sea ice drift product from Sentinel-1 SAR of the CMEMS cannot be ignored here. Please mention it, for example referring to the Copernicus website (http:marine.copernicus.euweb69-interactive- catalogue.php?option=com_csw&view=details&product_id=...).

The sentence has been changed to include the CMEMS data set: ***This introduces a new era in SAR earth observation. Sea ice drift data with medium resolution (10 km) is provided operationally via the Copernicus Marine Environment Monitoring Service (http://marine.copernicus.eu), but no sea ice drift algorithm using Sentinel-1 data has been published so far.***

In addition, CMEMS data has been used for validation of the ORB algorithm.

*) same page line 26: see discussion item above: the notion of independence is not well-defined here, and you cannot say that pattern-matching methods automatically lead to dependency, nor that feature-tracking implies the vectors are independent. Please discuss these aspects in the manuscript.

The sentence has been changed to: ***Current pattern matching algorithms constrain the high resolution vectors with low resolution estimates for practical reasons. Using feature tracking, drift vectors can be derived independently from the surrounding motion, which leads to better performance e.g. along shear zones.***

Section 5 is extended by a comparison between feature tracking and pattern matching with regard to independence.

*) It is customary to use the introduction to give a short layout of the rest of the paper. Please add this.

The following has been added to the introduction: ***The paper is organised as follows: Section 2 introduces the used Sentinel-1A data product. The ORB algorithm description and the used methods for tuning, comparison and validation are presented in Section 3. The recommended parameter set including the tuning, comparison and validation results are provided in Section 4. The discussion can be found in Sect. 5.***

*) section 2: Data. Is the Sentinel-1 mission an initiative of the sole ESA? The Sentinels are part of the boarder Copernicus, which is an EU programme. A joint initiative of ESA and EU?

The sentence has been changed to: ***The Sentinel-1 mission, an initiative of the European Union and operated by the European Space Agency (ESA), is composed of ...***

*) page 6940, lines 10-15. Mentions what is actually measured: radar backscatter sigma0.

The sentence has been changed to: ***Sentinel-1 carries a single C-band Synthetic Aperture Radar (SAR) instrument measuring radar backscatter at a centre frequency of 5.405 GHz and supporting dual polarisation (HH+HV, VV+VH).***

*) same page, line 24: capitalize 'Earth'.

Agree, has been changed.

*) page 6940 and 6941 The last paragraph were the Nansat tool is discussed and its handling of the GCPs is not part of 'data' and could be moved to 'methodology'.

The last paragraph of Section 2 has been moved to Section 3

*) Same paragraph. It sounds like you did little pre-processing of the data. Did you applying calibration, de-banding, transformation to dB scale, etc...? If yes, please specify these steps. If not, state clearly that the images are used as found in the product files.

Only radiometric calibration of the normalised radar cross-section (σ^0) was performed. This has been described in Section 3.

*) Page 6941 line 12: 'major' -> 'main'

Agree, has been changed.

*) Same page, line 23: mention that the parameters are the 'best' for your region of interest. This is elsewhere in the text. Your parameters are optimised for your region of interest. They are also optimised for winter/spring cases only.

The sentence has been changed to: ***The best suitable parameter set (including spatial resolution of SAR image, patch size of FAST descriptor, number of pyramid levels, scale factor etc.) has been evaluated for our area and season of interest.***

We mention the focus on winter/spring data in Section 2.

*) Page 6943, line 4: 9 contiguous pixelS. Next line, fix typo in 'recognized'.

Agree, has been changed.

*) Page 6943: while describing the algorithm, you seemingly use 'keypoint' and 'feature' for the same thing. Are they the same things and if not can you add a sentence to state the difference?

Keypoint and feature are not the same thing. Keypoint refers to a single pixel in the centre of a patch p that is used for feature description. Feature refers to the description of p. That is in our case the binary vector f. We tried to be more consistent about it throughout the manuscript and added: ***(NB: keypoint refers to 1 pixel, feature refers to description of*** p***)***

*) Pages 6943 6944: the reader might be lost by your steady pace at describing the ORB. Could you add a figure/sketch showing the location of the 256 pairs designed by Rublee et al. 2011? Or are they at random? Could you also give an example of computing the Hamming distance of 2 features? What is the 'ratio test from Lowe 2004' can it be described here in few sentences, including its purpose in your setting? Is it for discarding rogue vectors?

The ORB algorithm description have been revised and many details have been added. A figure showing a keypoint has been added, including the surrounding patch p used for feature description and an example for a binary test using 5x5 sub-windows. The Hamming distance and Lowe ratio test have been explained in more detail, including an example.

*) Page 6945: here some information about the ratio test. Please move this to the description of the method and only keep the discussion of the 0.75 value if still relevant.

The describing part has been moved to Section 3 and the rest has been kept.

*) Page 6945 lines 8-9: ', if the motion vectors are too different' (instead of 'if the matches are not correlated').

Agree, has been changed.

*) Page 6945 line 10: The reader can probably do the math, but it looks wrong to add km and km/h. Please fix the equation by introducing the time separation between the images.

The sentence has been changed to: ***Unreasonably high sea ice displacements (e.g. above 40 km for a time difference between two scenes of ~30 hours) are removed in a post-processing step from the drift field.***

*) Page 6945 line 15: 'computationally'. Later: 'best suitable' -> 'optimized'

Agree, has been changed.

*) Page 6945: The RMSD as a quality measure is dubious as it would favour algorithms that smooth a true discontinuous motion field (shear, divergences,...) It seems you are using it as a gross error check to detect outliers more than a ranking. Maybe discuss this?

The term 'quality measure' is not used anymore. The section has been renamed and split into 'Comparison of ORB to SIFT and SURF' and 'Validation'. More details have been provided on the root mean square error/distance calculation. See also General comment 4.5 .

*) Page 6946 line 6: 'velocities' are not in 'km'. You could use 'displacement' or 'motion'.

'velocities' has been replaced with '*displacement*'.

*) Page 6946 'Patch size'. Is Patch Size the n from Eq 3? Please make sure the reader knows what you are varying and how it relates to the description of the algorithm.

n **refers to the number of binary tests. In case of ORB** $n = 256$**. The patch size refers to the size of the considered patch** p **around the keypoint used for feature description. The revised algorithm description including an example of a keypoint incl. patch** p **should make it more clear how the varied parameters relate to the algorithm. The sentence has been change to:** ***Figure 2 shows that changing the size (length and width) of the considered patch*** p ***between 10 and 60 pixel can modify the resulting amount of vectors by an order of magnitude.***

*) Section 4.1.2 : Refer the reader back to Eq 1. Are the values you find here expected to work during summer? Please briefly discuss the expected adequacy of your values to all-year-round (in this case: summer) applications. This discussion could be moved to Section 5 'Discussion and Outlook'.

A reference to the Equation has been added. The discussion has been extended considering summer/autumn data.

*) Section 4.1.3 : It seems features are tracked at all pyramid levels. So adding a pyramid level logically brings more features. Is it what we see in Figure ?

Adding a pyramid level, does not necessary bring more matches. Note that the numbers of matched features decreases slightly towards 14 pyramid levels. This is now better visible, since the four image pairs have been shown separately. The fixed parameter is 100 000 keypoints. These keypoints are distributed over all pyramid levels, hence more pyramid levels might bring less keypoints in the low (fine scale) pyramid levels. The last sentence has been changed to: ***As shown in Figure 4, the number of matches does not increase significantly when using more than 7 pyramid levels and even decreases towards 14 pyramid levels.***

*) section 4.1.4 : Again, please specify that the coefficients in table 2 are optimized for your region of interest.

The sentence has been change to: ***Table 2 shows the recommended parameter set for ORB Sentinel-1 sea ice drift application for our region and period of interest.***

*) section 4.2 : This is where this paper should try and give more insights as to why HV channel gives so many more vectors than the HH channel. Are they located at the same locations? A figure showing a (zoomed) Sentinel-1 SAR image and the location of HV and HH features is suggested (if it serves its purpose).

See General comment 1.3 .

*) section 4.3 : It seems that the authors only tuned the parameters of the ORB algorithm, while the other two (SURF and SIFT) are used 'out-of-the-box', without optimizing. If the case, it is not surprising to find better results for ORB (after optimization). Please modify the text (also abstract and conclusion) so that it is clear to the reader which algorithms were optimized (and why your choice is still ORB - for example license issues and computational efficiency).

A Section 'Comparison of ORB to SIFT and SURF' has been added to the methods, explaining the chosen set-up of SIFT and SURF and why ORB has been used for tuning.

*) section 4.5 validation. The RMSD to the manual vectors is not impressive (600m) considering that they are from the same images. Komarov and Barber (2014) find RMSD of about 400m against GPS drifters, and similar values are reported for the CMEMS/DTU Sentinel-1 based product. To be more conclusive, please load the CMEMS 10-km Sentinel-1 SAR product (from DTU) computed from this pair of Sentinel-1 images, and do the validation against the compute the RMSD to this product. The validation will then not suffer from the uncertainty due to manual matching.

The validation has been extended to include CMEMS data. Three comparisons have been made: ORB vs manual, ORB vs CMEMS and CMEMS vs manual for comprehensive assessment of the algorithms. Root mean square error (E) of ORB vs manual is 563 m, whereas E of CMEMS vs manual is 1690 m indicating that the feature tracking approach outperforms pattern matching. The validation methodology has been described in Section 3.4 and the results have been presented in Section 4.5, Figure 5 and Table 4.

*) page 6949: Discussion and Outlook. This section is too short and more discussions are needed for this manuscript to be more than a report about your investigations. In the revised manuscript, it should be extended with the discussion of the differing HV/HH performances, and the discussion of the distance between features compared to the size of the features (are the motion vectors from two neighbouring features independent?). This will then be a very good contribution to this journal.

The discussion has been extended considering error sources, comparison of feature tracking and pattern matching with regard to 'independence' and resolution, applicability of the ORB algorithm and HH versus HV polarisation performance.

*) Your Outlook is too vague. Rather state what your next steps will be in your investigations of Sentinel-1 sea ice drift.

A more specific outlook has been added.

*) Figures 1, 2, 3: it is puzzling to see the HH+HV (red) curve in-between the HH and HV (black) curves. Should not the red curve be the sum of the other two?

Agree, has been changed.

2 Response to anonymous referee #2

General comment

I find the descriptions of the algorithm to be insufficient for understanding the merit of the procedure. While most readers are not experts in motion tracking, the authors should provide at least enough details for a general appreciation of the merits of ORB. Not enough was provided in the approaches of SIFT and SURF, and it was not clear whether these algorithms were tuned to obtain the best results. Also, this is a small data set to be used for quality assessment of these algorithms. Thus. the conclusion seems to be somewhat unjustified given the restricted data set used in the assessment. A proper evaluation of these algorithms, seems to me, requires a larger data set with better-defined metrics for objective analyses.

The ORB algorithm description in Section 3.1 has been revised and more details are given for a better understanding of the procedure. A new figure (Figure 1) has been added, displaying a recognised keypoint of the 'Fram Strait' image pair including its orientation θ and the considered patch p around the keypoint that is used for feature description. SIFT and SURF were used in standard mode and the framework conditions were set equal for the comparison. That means, the image pre-processing has been done in the same way, Brute Force Matching including a Lowe ratio test with 0.75 has been applied for all three algorithms, as well as the removal of unreasonably high sea ice displacements in a post-processing step. Since SIFT allows to define the number of returned keypoints, this parameter has been set to 100 000 as done for ORB. The further tuning of SIFT and SURF is not the aim of this paper, since these two algorithms are not open source and computationally less efficient.

Specific comments

*) Page:line 6940:20 Perhaps it would be useful to describe why the selected images are sufficient for evaluation process? i.e., the range of ice conditions and what seasons do they cover? It seems that the data covers only winter conditions.

The following has been added to the data description: ***We focused on winter/spring data, since our area of interest experiences the highest sea ice cover during this period.***

The following has been added to the discussion: ***The algorithm tuning has been performed using winter/spring data, since our area of interest experiences the highest sea ice cover during this period. During summer/autumn, most considered areas have very little or no ice cover (e.g. Barents Sea and Kara Sea), making ice drift calculation during this period less meaningful. Nevertheless, some areas, like western Fram Strait, experience sea ice cover during the entire year. Dependence of the algorithm performance on the season needs to be evaluated in future work. Computing sea ice drift from summer/autumn data is expected to be more demanding, since features might be destroyed by melting.***

*) 6941:5 What specific geometric coordinate system are the images projected onto prior to motion tracking? Are these ground range images? If so, what are residual planimetric distortions that could contribute to the assessment of data quality?

The original images are not projected prior to motion tracking in order to prevent distortion of features by data resampling during the reprojection. Yes, these are ground range (GRD) images. The reported distortions of raw Sentinel 1A SAR data (Schubert et al., 2014) is within 10 m and contributes to the error budget only very slightly. The following has been added to the data description: ***...residual planimetric distortions: within 10 m (Schubert et al., 2014)...***

*) 6941:15 The resampling algorithms listed do not necessarily reduce speckle noise, e.g., nearest neighbor, bilinear, etc. It is not surprising that simple averaging was the best.

Part (a) has been changed to: ***To decrease the influence of speckle noise and increase the computational efficiency, the resolution is reduced before applying the ice drift algorithm from 40 m to 80 m pixel spacing using simple averaging.***

*) 6942:1 Define grid cell.

Part (c) has been changed to: ***The introduced ORB setup is compared to other available OpenCV feature tracking algorithms, CMEMS data and manually drawn vectors for performance appraisal and validation.***

*) 6943:1- It would be useful if the authors provided remarks along the way to help the reader follow the construction of features based on image intensity. How important is the absolute/relative calibration of the image values to ORB? So, the binary keypoints are no sensitive but the score R is?

The ORB algorithm description in Section 3.1 has been revised and more details are given for a better understanding of the procedure. Keypoint detection and feature description are based on binary tests which are not sensitive to image calibration. A more detailed description for R including equation has been added. R depends on the intensity variation of a surrounding patch and is therefore somewhat sensitive to image calibration. However, R is only used as relative value to filter out less reliable features. The importance of image calibration has been considered in three ways: (1) Converting the linear backscatter values before the transformation into decibel units has been tested, but decreased the algorithm performance for both HH and HV channel and, therefore, was omitted. (2) A range of values for minimum and maximum backscatter (Equation 2) have been tested. The range of the individual lines in Figure 3 give an impression on how image calibration influences the algorithm performance. (3) The following has been added the the discussion: ***Using noise removal for HV and angular correction for HH has been tested, but did not improve the feature tracking results, i.e. a lower number of vectors has been found.***

*) 6944:10 What does the feature vector actually look like? How many features are there? It's not clear in the text. How are these features normalized for distance calculations?

The ORB algorithm description in Section 3.1 has been revised and more details are given for a better understanding of the procedure. For each found keypoint, a feature vector is calculated based on the patch p around the keypoint, i.e. number of features = number of keypoints (100 000 per image, in best case). Each feature vector consists of $n = 256$ binary values (0 or 1). The vectors are not normalised. 'Distance' in this context refers to Hamming distance. An example for Hamming distance has been added to the algorithm description.

*) 6945:10 Is it really per hour?

The sentence has been changed to: ***Unreasonably high sea ice displacements (e.g. above 40 km for a time difference between two scenes of ~30 hours) are removed in a post-processing step from the drift field.***

*) 6945:12 I don't know how this could be justified: '...Using this assumption, it can be concluded that more matches equals better algorithm performance...'

The sentence has been changed to: ***Based on our observations we assume that the proportion of wrong matches does not increase with increasing total number of matches. Under this assumption the algorithm performance refers to the total number of matches and is used to tune the algorithm parameters in Table 2.***

*) 6946: 1-10 Should the reader be familiar with these parameters?

The revised ORB algorithm description in Section 3.1 should give a better understanding of the tuned parameters. In addition, the sentence has been changed to the following: ***As a starting point, the tested parameters were set as follows: resize factor = 0.5, patch size = 31, pyramid levels = 8, scale factor = 1.2, HH limits = [0,0.12], HV limits = [0,0.012] and ratio test = 0.8. As a compromise between performance and computational efficiency, the amount of maximum retained keypoints is set to 100 000. Tested range and parameter meaning are shown in Table 2.***

*) 6946: 4 2.5 km is not a velocity. Do you mean 2.5 km per day?

The sentences have been changed to the following: ***In addition, displacements below 2.5 km are rejected during the testing to disregard matches over land. This does not influence the number of correct matches, since the sea ice displacement in all considered test images is above 2.5 km.***

*) 6947:3 What does the scale factor do?

The following has been added to the revised algorithm description in Section 3.1: ***A scale factor of 2 means that each next pyramid level has 4 times less pixel, but such a large scale factor degrades the feature matching score. On the other hand, a small scale factor close to 1 means to cover a certain scale range needs more pyramid levels and hence, the computational cost increases.***

*) 6947: How are the SIFT and SURF algorithms tuned?

The following has been added to the method description: ***SIFT and SURF were used in standard mode and the framework conditions were set equal for the comparison. Image pre-processing has been done as described above, Brute Force Matching including the Lowe ratio test with threshold 0.75 has been applied for all three algorithms as well as the removal of unreasonably high sea ice displacements in a post-processing step. Since SIFT allows to define the number of retained keypoints, this parameter has been set to 100 000 as done for ORB. The further tuning of SIFT and SURF is not the aim of this paper, since these two algorithms are not open source and computationally less efficient.***

*) 6948:19 In general, how far are the vectors from the manually identified vectors? Should the quality be dependent on distance from manual observation?? What is the RMSD for the other two algorithms?

The maximum distance between validation and reference vector was set to be 5 km. Average distance between validation and reference vectors as well as root mean square error, slope and offset of the linear fit are added to the validation results and shown in Table 4. The tuned ORB algorithm has been compared to SIFT and SURF using number of matches N and root mean square distance D (see revised method description). Since the tuned ORB algorithm

outperforms SIFT and SURF and the further use of SIFT and SURF is not considered, we did not validate these two algorithms against the manual derived drift vectors. However, CMEMS ice drift data has been added for further validation and comparison against the manual derived drift vectors.

3 Response to anonymous referee #3

General comments

3.1

The method should be described in some more detail, preferably with a figure explaining the feature identification procedure.

The ORB algorithm description in Section 3.1 has been revised and more details are given for a better understanding of the procedure. A new figure (Figure 1) has been added, displaying a recognised keypoint of the 'Fram Strait' image pair including its orientation θ and the considered patch p around the keypoint that is used for feature description. See General Comment 2.

3.2

The results from (only) 4 example images are presented very summarically. It should be shown how the results vary between the 4 examples (preferably by showing results from all 4).

Figure 2, 3 and 4 have been changed to show the respective results of the 4 considered image pairs. Corresponding result description and discussion have been added to Section 4 and 5.

3.3

The discussion section should include a further discussion (based on the differences between the 4 examples) on the applicability of the algorithm in other regions and seasons than the examples (which are all from March-April). In particular it should be discussed in more detail how different ice type regimes would impact the results.

The discussion has been extended based on the differences of the 4 example image pairs. The region and time period of interest has been considered, as well as the performance in different ice conditions.

Specific comments

*) P6939-L20-21: Even though a sea ice drift algorithm specifically for S1 may not have been published, the S1 data are very similar to Radarsat data and to some extent also Envisat ASAR data, and the Copernicus Marine Service operationally produce daily ice drift data from Sentinel-1 data, so in that respect the results should have been compared to the CMEMS ice drift data available from the CMEMS data portal. The CMEMS ice drift data are from day to day so the first two example datasets should have CMEMS correspondence.

The CMEMS/DTU data set has been included and the vector field provided for the image pair 'Fram Strait' has been utilised. CMEMS/DTU data have been compared to ORB vectors and the presented manual derived drift vectors, as shown in Section 3.4 and 4.5, Figure 1 and Table 1.

*) P6940-L7: The Sentinel-1 satellite belongs to the Copernicus programme of the European Union and should be referred to as such, NOT as an ESA satellite. ESA operates the satellite but Copernicus is an initiative of the European Union.

The sentence has been changed to: ***The Sentinel-1 mission, an initiative of the European Union and operated by the European Space Agency (ESA), is composed of ...***

*) P6940-L18: The resolution of the EWS GRDM S1 data is 93x87 meters with an ENL of 12.7 according to https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/resolutions/level-1-ground-range-detected The pixel spacing is 40x40 meters but that is not the same as resolution.

The sentence has been changed to: ***These images cover an area of 400×400 km with a pixel spacing of 40×40 m (resolution: 93 m range $\times$ 87 m azimuth) and provide both HH and HV polarisation.***

*) P6941-L19: Same issue as above. The resolution is not 40x40m in the S1 EWS data. Please rephrase.

The sentence has been changed to: ***Best performance and computational efficiency was achieved by using simple averaging from 40 m to 80 m pixel spacing.***

*) P6942-L20-25. Please clarify if this transformation is universal or performed patch by patch, image by image or what? Sigma-max and sigma-min are user defined as far as I understand, but that should be stated clearly in order to avoid the misunderstanding that they depend on image or patch properties (and thereby speckle noise).

The following sentence has been added after Equation 2: ***Lower and upper brightness boundaries $\sigma^0_{\min}$ and $\sigma^0_{\max}$ are user defined and chosen to be constant in order to limit the influence of speckle noise and be independent of e.g. high backscatter values σ^0 over land.***

*) P6943-L8: addressed -> ascribed

The Harris corner measure description has been changed and does not include this phrase anymore.

*) P6945-L15-16: changed to 'computationally more efficient enabling testing the remaining parameters'

Agree, has been changed.

*) P6945-L24: what is a grid cell?

The comparison to SIFT and SURF has been described in more detail and grid cell has been defined in the following sentence: ***The distribution and reliability of the calculated vector fields have been assessed using two parameters on a grid with cell size 1 $^\circ$Longitude $\times$ 0.2 $^\circ$Latitude: number of derived vectors per grid cell (N) and root mean square distance (D) of all vectors in a gird cell computed as follows: ...***

*) P6945-L20-25: It should be discussed if individual features tracked can be considered independent.

See general comment 1.2 .

*) P6947-L23: The figure only shows 1.2 (not 1.1, 1.3 and 1.4)

Agree, has been changed.

*) P6947-L20-24: The ORB algorithm was optimized before this inter-comparison. It should be more clearly stated whether the SIFT and SURF algorithms were also optimized (or if they were not, some considerations on their potential optimization would be useful).

The following has been added to the method description: ***SIFT and SURF were used in standard mode and the framework conditions were set equal for the comparison. Image pre-processing has been done as described above, Brute Force Matching including the Lowe ratio test with threshold 0.75 has been applied for all three algorithms as well as the removal of unreasonably high sea ice displacements in a post-processing step. Since SIFT allows to define the number of retained keypoints, this parameter has been set to 100 000 as done for ORB. The further tuning of SIFT and SURF is not the aim of this paper, since these two algorithms are not open source and computationally less efficient.***

*) P6948-L1-5: It should be discussed whether the vectors found with the ORB algorithm can be considered independent. This is a very important prerequisite of the conclusion here, and it is

not clear whether the many vectors are not partly dependent. SIFT and SURF algorithms can also provide many more vectors if they are applied with many more partly overlapping windows.

See general comment 1.2. The further tuning of SIFT and SURF is not the aim of this paper, since these two algorithms are not open source and computationally less efficient. SIFT and SURF can be tuned to obtain a higher number of vectors, but considering that SIFT and SURF provide less vectors accompanied by higher root mean square distance D values, suggests that increasing the number of matches will result in a higher number of wrong matches compared to the tuned ORB algorithm.

*) P6948-L7: are sufficient computational efficient (not good english)

The sentence has been changed to: ***The OpenCV feature tracking algorithms ORB, SIFT and SURF in combination with the python-toolbox 'Nansat' are computationally efficient (total processing time on regular MacBook Pro: 2–4 minutes) and allow high resolution sea ice drift retrieval from datasets with large temporal and spatial extent.***

*) P6948-L15-21: The validation illustrated in figure 4a show results for all 3 algorithms. Numbers should be provided for RMSD between the 3 algorithms and between the 3 algorithms and CMEMS operational ice drift data. 7.5 pixels (610 meters) seems like quite a large difference, so some additional information about whether the difference is randomly distributed of mainly present in areas of ice rotation or other characteristics would also be relevant.

The validation has been extended and CMEMS data has been included. Three comparisons have been made: ORB vs manual, ORB vs CMEMS and CMEMS vs manual for a comprehensive assessment of the tuned ORB algorithm. Root mean square error E of ORB vs manual is 563 m, whereas E of CMEMS vs manual is 1690 m, indicating that the feature tracking approach outperforms pattern matching. The range of displacement lengths is 10–35 km. Hence, the relative error of the algorithm is very low (about 2.5 %). The validation method is described in section 3.4, the results are presented in Section 4.5. and illustrated by figures and tables.

*) P6949-L4: proof -> prove

Agree, has been changed.

*) P6949-L6: proofing - proving

This part has been rewritten and does not include proofing anymore.

*) P6949-L8-11: This is speculation and should be demonstrated with data. It has not been proven that feature tracking vectors are necessarily independent, and it has not been shown that pattern recognition is more prone to errors in areas of high velocity gradients. Evidence should be provided.

Section 5 is extended by a comparison between feature tracking and pattern matching with regard to independence and resolution. Two Figures have been added showing sea ice drift anomalies on a zoomed sub-image and a variogram of drift vectors.

*) P6950-L6: Synthetical - Synthetic

The name of the project is spelled with 'Synthetical'.

*) Figure 1, 2 and 3: Please show results for each of the 4 test cases and in the text discuss what may cause different results in the 4 cases.

Figure 2, 3 and 4 have been changed to show the respective results of the 4 considered image pairs. Corresponding result description and discussion have been added to Section 4 and 5.

4 Response to referee Tim Dunker

General comments

4.1

I think the writing needs an overhaul: there are a number of spelling mistakes and grammatical errors, and physical terms and units are not used correctly. I think the title should be rearranged. Some statements are ambiguous, because you use a mix of active and passive voice. The validation (Sect. 4.5) and discussion (Sect. 5) parts are very short. There is, in my opinion, hardly any discussion. The headline mentions an outlook, but I do not think there is enough there to justify this headline. Only the very last sentence of that section might be an outlook, but because you use passive voice, readers will not know if you recommend the combination of methods or if this is some general statement. An improved structure of discussion and conclusion (and outlook, if applicable) will not require too much work, but will help the reader to better navigate through the manuscript. I offer a number of specific comments and suggestions you may or may not consider. Hopefully, you will find some of them helpful.

The manuscript has been revised and several spelling mistakes and grammatical errors have been corrected. We paid attention during the correction to use all physical terms in a correct manner. The title has been rearranged. The validation and discussion part have been extended and the structure has been rearranged.

4.2

Though Fig. 4(c) supports your claim that the adapted ORB algorithm yields superior performance to SIFT and SURF, it would be instructive to see this comparison for a range of parameter choices (see Table 2). Furthermore, instead of showing an image of results in Figs. 4(b,c), the actual distribution function would reveal many more details of the algorithm's performance, I believe. Especially if there were a potential bias in any of the three algorithms towards large or small vector norms, this bias would stand out clearly in a plot of the distribution function or a quantile-quantile plot, for instance.

We did not vary the parameters for the comparison, since SIFT and SURF do not have the same tuning parameters as ORB. They were used in standard mode and the framework conditions were set equal for the comparison. The further tuning of SIFT and SURF is not the aim of this paper, since these two algorithms are not open source and computational less efficient (see General Comment 2). A quantile-quantile plot (Figure 9) comparing the distributions of number of vectors (N) and root mean square distance (D) has been added to Section 4.3. It supports our findings that N of the ORB algorithm is in all cases higher than N of SIFT and SURF, and that D of ORB is (almost) always lower. D of ORB is slightly higher in the very low range, due to a higher number of vectors N of the ORB algorithm results.

4.3

Figures 1, 2, and 3 show average results from the four regions mentioned in Table 1. From these figures, it is not possible to judge where the algorithm performed better or worse, as indicated by the number of matches. Instead of having one aggregate figure, would it be better to have one panel per region per figure? In my opinion, this would clarify which parameter should be chosen for which sea ice condition. I expect the sea ice conditions in the Kara Sea and around Franz Josef Land to be quite different compared with the Fram Strait - namely, much less and slower motion in the Kara Sea and maybe more ridges around Franz Josef Land. I think these different conditions make possible a nice comparison of your algorithm with SIFT and SURF.

Figure 4, 5 and 6 have been changed to show the respective results of the 4 considered image pairs. The discussion has been extended based on the differences of the 4 example image pairs and the performance in different ice conditions has been considered (see general comments 3.2 and 3.3).

4.4

I am curious to know how the algorithm performs if there are melt ponds on the sea ice surface. Unfortunately, the four test image pairs are all from March and April, when typically the sea ice is largely snow-covered and melt ponds are not as common as in late spring or summer. Therefore, I contest your statement that these image pairs are 'representative [of your] region of interest' (p. 6940, l. 19; see also the Abstract, ll. 3 to 5). These images might be representative of the sea ice conditions mentioned in the same sentence (p. 6940, ll. 20 to 21), though.

The algorithm tuning has been performed using winter/spring data, since our area of interest experiences the highest sea ice cover during this period. Nevertheless, some areas, like western Fram Strait, experience sea ice cover during the entire year. Future work needs to evaluate the algorithm performance depending on the season. Computing sea ice drift from summer/autumn data is expected to be more demanding, since features might be destroyed by melting. Section 2 and Section 5 have been extended to mention and discuss the choice of the considered four image pairs in more detail. The open-source distribution of our algorithm shall ensure that everyone can apply the algorithm in their area/condition of interest.

4.5

I do not think I fully understand the quality measure and how you applied it, maybe I have just misunderstood it. In Sect. 4.5 you describe the RMSD: '... combination of the manually produced error and the displacement variation between the manual and calculated vector.' In Figure 4(c), you show the RMSD for SIFT and SURF, but you do not mention any manually-drawn vectors that you might have compared to SIFT or SURF results. How did you compute the RMSD for SIFT and SURF? What do you mean by 'manually produced error'? You also compute the numbers of correct and wrong matches, but you do not use these to assess quality. Are these figures not be appropriate for qualitiy assessment? Given that the manuscript is about sea ice velocity, should the quality measure not be a velocity either? Have you attempted to quantify the uncertainty of the drift speed?

The term 'quality measure' is not used anymore. The section is renamed and split into 'Comparison of ORB to SIFT and SURF' and 'Validation'. More details are provided on the root mean square error/distance calculation. The manual produced error is simply the one introduced by the sea ice expert. This is now mentioned in more detail in the Discussion. We observed that proportion of wrong matches is nearly constant if we vary/test certain parameters like the patch size and, therefore, did not include these figures. We consider only displacement of sea ice between two image pairs taken at a given time difference (whereas the actual sea ice velocity can be higher). The uncertainty of the displacement vectors is quanitified by comparing the results of image pair 'Fram Strait' with manually derived vectors. Our validation shows that the error of the tuned ORB algorithm is expected to be around 563 m.

4.6

Before publication, this comment must be addressed: Parts of Section 3.1 are a basically a repetition with negligible alterations of Section 4.1 in Rublee et al. (2011), most notably page 6943, line 19, to page 6944, line 8. One might get the impression that the mathematics is part of your work. However, the mathematics is exactly as published by Rublee et al. (2011). If I am not mistaken, Rublee et al. (2011) already used brute-force matching and the Hamming distance, so Rublee et al. (2011) should also be cited on page 6942, lines 16 to 17. There are some direct quotes (p. 6942: ll. 7 to 9; ll. 14 to 16; ll. 17 to 18) of Rublee et al. (2011) with minimal changes, but are not marked as such. In the last of these sentences, you have changed the meaning of the sentence, because you have substituted 'of' for 'unlike'. Any direct quote must be enclosed by quotation marks, and modifications must be clearly indicated. Either use quotation marks or reword entirely, but always cite Rublee et al. (2011). Another solution could be to drop most of Section 3.1 as it is now, and instead cite Rublee et al. (2011) for the ORB algorithm and concentrate on the adaptations you have made.

We included an ORB algorithm description, similar as shown in Rublee et al. (2011), since this feature tracking method is new to the sea ice drift community. Anonymous Referee #1, #2 and #3 asked for an even more detailed algorithm description. Section 3.1. has been revised and edited accordingly taking correct citation of Rublee et al. (2011) into account. Even though Rublee et al. (2011) uses Brute Force matching and Hamming distance for their result evaluation, these methods are not part of the ORB algorithm itself. OpenCV allows different matching procedures for ORB. We found the best results for our set up with Brute Force matching and Hamming distance. Nevertheless, we mentioned that Rublee et al. (2011) also uses Brute Force matching and Hamming distance.

Specific comments

*) In most instances throughout the manuscript, 'computational' must be an adverb, not an adjective. On page 6941 (lines 9, 15, and 18), page 6944 (line 18), page 6947 (lines 4 and 9), and on page 6948 (line 6), the use of an adjective is correct. Please search for 'computational' and replace it by 'computationally', where appropriate. Please also replace 'pixel' with 'pixels', where appropriate. In general, I vote for the use of basic SI units, such as ms^{-1} instead of kmh^{-1}, though I see that the latter appears more practical when two radar images are separated by many hours or days.

'Computational' and 'pixel' have been replaced by 'computationally' and 'pixels', where appropriate. Since we consider the displacement of sea ice between two images, we use now only km rather than velocity units.

*) Title: You may want to think about rearranging the title. Now, readers might be led to believe the manuscript is mostly about sea ice drift estimated from the feature tracking method. Instead, the manuscript is about the feature tracking method applied to sea ice drift detection. Would not 'An open source feature tracking algorithm to detect sea ice drift from Sentinel-1 SAR imagery' be more suitable? I think so.

The title has been changed to: ***Open source feature tracking algorithm for sea ice drift retrieval from Sentinel-1 SAR imagery***

*) Page 6938, lines 3 to 5: I do not think it is justified to call the image pairs simply 'representative'. They might be representative of the sea ice conditions you investigate, but not necessarily representative in general. Please specify what you consider the images to be representative of.

The sentence has been changed to: ***The best suitable setting and parameter values have been found using four Sentinel-1 image pairs representative of sea ice conditions between Greenland and Severnaya Zemlya during winter/spring.***

*) Page 6938, lines 5 to 6: Comma after 'introduced'. I generally prefer writing 'use' instead of 'utilise' (also on page 6941, line 7; page 6941, line 22; page 6943, line 11; page 6949, line 13), unless there is an urgent need for exactly that word. Besides, I do not think that this sentence is strictly necessary, and I think you overstate the quality measure's novelty and importance a little. From lines 11 to 13, page 6938, it is not clear that this refers to the quality measure.

The sentence has been removed and the term 'quality measure' is not used anymore. 'Utilise' has been replaced by 'use' throughout the manuscript.

*) Page 6938, line 10: Decimal in '2.7 GHz'

Agree, has been changed.

*) Page 6938, line 12: Comma after 'vectors'

Agree, has been changed.

*) Page 6938, line 13: 'significantly'

Agree, has been changed.

*) Page 6938, line 14: 'four times'

Agree, has been changed.

*) Page 6938, line 18: I propose to change 'observe from earth observation data' to 'observe from remote sensing data'.

Agree, has been changed.

*) Page 6938, lines 23 to 24: I disagree. A little bit further into Sect. 1 (page 6939, lines 9 to 12), you write the opposite: data with a spatial resolution of 400m have been available for some years. Maybe you should clarify what you mean by 'sufficient resolution'.

The sentence has been replaced by: ***However, there is still a lack of large sea ice drift data sets with sufficient resolution to estimate convergence and divergence on a spatial scaling of less than 5 kilometres.***

*) Page 6938, lines 23 to 24: I do not think 'convergence field[s]' or 'divergence field[s]' is correct, because divergence is a scalar. The divergence operator is applied to a field, for example to a velocity field. I suggest to simply write 'divergence' and/or 'convergence'.

Agree, has been changed to: ***... convergence and divergence on ...***

*) Page 6938, line 25: 'The regions of interest are'.

Agree, has been changed.

*) Page 6939, line 1: 'These seas' or 'These areas'?

The sentence has been changed to: ***This area is characterised by ...***

*) Page 6939, line 8: 'which acquires data'

Agree, has been changed.

*) Page 6939, lines 18 to 19: One day is not a frequency.

The sentence has been changed to: ***... unprecedented revisit time of less than one day ...***

*) Page 6940, line 1: '. . . has been considered' by you or by Rublee et al. (2011)? Please clarify.

The sentence has been changed to: ***For application on large data sets and for operational use, we considered a computationally efficient algorithm, called ORB (Rublee et al., 2011), tuned it for sea ice drift retrieval from Sentinel-1 imagery and compared the results with other available feature tracking algorithms and existing sea ice drift products.***

*) Page 6940, line 18: Either 400 km × 400 km or (400 × 400) km^2 .

Has been changed to: ***400 km × 400 km***

*) Page 6940, line 13: One day is not a frequency.

The sentence has been changed to: ***With both satellites operating, the constellation will have a revisit time of less than 1 day in the Arctic.***

*) Page 6940, line 19: 'representative of '

Agree, has been changed.

*) Page 6940, line 23: You provide all relevant links in Appendix A already. I suggest to either move the link to a footnote here or write ’(see Appendix A; Korosov et al., 2015)’. With either option, you get rid of the two subsequent parentheses.

The reference has been changed to: ***(see Appendix A; Korosov et al. (2015, 2016))***

*) Page 6940, line 24: ’scientist-friendly’

Agree, has been changed.

*) Page 6940, line 26: ’a simple’

Agree, has been changed.

*) Page 6941, line 7: For consistency: ’centre’ and ’centred’

Agree, has been changed.

*) Page 6941, line 10: ’at high latitudes’

Agree, has been changed.

*) Page 6941, lines 13 to 14: I think you overstate a little here regarding the quality measure.

The term ’quality measure’ is not used anymore. The sentence has been changed to: ***Our algorithm for sea ice drift detection includes three main steps: (a) resampling of raw data to lower resolution, (b) detection and matching of features and (c) comparison/validation.***

*) Page 6942, lines 1 to 2: Commas after ’measure’ and ’cell’.

The sentence has been replaced by: ***The introduced ORB setup is tested against other available OpenCV feature tracking algorithms for comparison, and CMENS data and manually drawn vectors for validation.***

*) Page 6942, line 11: ’applies a Harris’

Agree, has been changed.

*) Page 6942, lines 11 to 12: What does ’them’ refer to - pyramid levels or multi-scale features? Your description, specifically the order of steps, seems to be different from the one given by Rublee et al. (2011). Please clarify.

The sentence has been changed to: ***It uses FAST to find multiscale-keypoints on several pyramid levels and applies a Harris corner measure (Harris and Stephens, 1988) to pick the best keypoints.***

The ORB algorithm description in Section 3.1 has been revised and the order of steps has

been adjusted.

*) Page 6942, line 19: ’applied to’

Agree, has been changed.

*) Page 6942, line 20: The variable σ^0 does not show up in Eq. (1). I suppose σ in Eq. (1) should read σ^0. I am not exactly sure if I understand the formulation ’i range [0,255]’ correctly. I think you mean that the intensity $i \in \mathbb{R}$ is defined on the interval [0,255]. If so, you could write something like $0 \leq i \leq 255$ for $i \in \mathbb{R}$.

σ has been replaced by σ^0 throughout the manuscript. The sentence has been changed to: ***Before the feature tracking algorithm can be applied to a satellite image, the SAR backscatter values σ^0 have to be transformed into the intensity i range ($0 \leq i \leq 255$ for $i \in \mathbb{R}$) used in openCV.***

*) Page 6942, lines 25 to 26: ’for both channels.’

Agree, has been changed.

*) Page 6943, lines 3 to 4: ’with a perimeter’

Agree, has been changed.

*) Page 6943, line 4: ’nine contiguous pixels’

Agree, has been changed.

*) Page 6943, line 5: ’recognised’

Agree, has been changed.

*) Page 6943, line 6: ’as a keypoint’

Agree, has been changed.

*) Page 6943, line 7: What do you mean by ’retained keypoints’?

The sentence has been changed to: ***The threshold t is set low enough to get more than the predefined amount N of keypoints.***

*) Page 6943, line 8: ’keypoint is assigned’

The corresponding sentence has been replaced.

*) Page 6943, line 9: Something went wrong with the citation here. I suppose you want a semicolon after ’Harris corner measure’.

The corresponding part has been replaced by: ***Harris corner measure (Harris and Stephens, 1988) is used to order the FAST keypoints ...***

*) Page 6943, line 16: ’intensity-weighted’

Agree, has been changed.

*) Page 6944, lines 8 to 9: Did you do the same as Rublee et al. (2011)? Because you use passive voice here, this statements does not say anything about what you did or did not do. Please clarify.

The set of sampling pairs was designed by Rublee et al. (2011) for the ORB algorithm. The sentence has been changed to: ***Rublee et al. (2011) applied a greedy search on a large training dataset to obtain a set for ORB with*** $n = 256$ ***relatively uncorrelated tests with high variance.***

*) Page 6944, line 19: ’40 m to 80 m’

Agree, has been changed.

*) Page 6945, line 2: No comma after parenthesis

Agree, has been changed.

*) Page 6945, line 8: No comma after ’rejected’

Agree, has been changed.

*) Page 6945, line 9: ’Unreasonably’; ’10 km + 1 kmh^{-1}’ is not a velocity.

The sentence has been changed to: ***Unreasonably high sea ice displacements (e.g. above 40 km for a time difference between two scenes of*** $\sim$***30 hours) are removed in a post-processing step from the drift field.***

*) Page 6945, line 15: ’equal’

Agree, has been changed.

*) Page 6946, line 1: This sentence confused me. You do not say which parameters you are going to test, so it is not clear to me what the ’remaining parameters’ are. At first, I understood this statement such that you did not test pyramid levels, scale factor, HH and HV limits, and ratio test, but set these to the mentioned values instead. From Table 2 and Figures 2 and 3, however, it becomes clear that you tested these settings over some range. Please clarify which parameters you tested and over what range (refer to Table 2, if appropriate) and which parameters you did not test.

The sentence has been changed to: ***As a starting point, the test parameters were set as follows: resize factor = 0.5, patch size = 31, pyramid levels = 8, scale factor = 1.2, HH limits = [0,0.12], HV***

limits = [0,0.012], ratio test = 0.8. Tested range and parameter meaning are shown in Table 2.

*) Page 6946, line 4: A 'low speed filter with 2.5 km' does not say much about the filter's nature, nor is 2.5 km a velocity. From what you write on the same page in lines 6 to 7, I first thought you used a high-pass speed filter. However, the filter rejects unreasonably high velocities, so I think you constructed a bandpass filter with a passband between your chosen cut-off speeds. Is this correct?

The sentences have been changed to the following: ***In addition, displacements below 2.5 km are rejected during the testing to disregard matches over land. This does not influence the number of correct matches, since the sea ice displacement in all considered test images is above 2.5 km.***

*) Page 6946, lines 6 to 7: 2.5 km is not a velocity.

'Velocity' has been replaced with 'displacement'.

*) Page 6946, line 13: 'Taking this'

Agree, has been changed.

*) Page 6946, line 15: I assume a pixel covers an area. Do you mean 2.72km^2 or (2.72 x 2.72) km^2 ?

The conversion has been removed since the size depends on the pyramid level.

*) Page 6946, line 16: Comma before and after 'respectively'

The sentence has been changed to: ***For our training dataset (Table 1), this yields on average around 1 and 4 vectors per 10 km^2 for HH and HV, respectively.***

*) Page 6946, lines 21 to 22: ' . . . image pairs, we suggest to set the maximum backscatter σ_{max} to 0.08 and 0.013 for HH and HV.'

Agree, has been changed.

*) Page 6946, lines 23 to 24: This statement ('decreasing towards higher values') is a bit confusing at first. To avoid misunderstandings, you could write something like '... , because the number of matches decreases for increasing values ofσ_{min} (not shown).'

Agree, has been changed.

*) Page 6947, line 1: '1 to 16'. Figure 3 only shows data for a scale factor of 1.2, but not for the other scale factors mentioned here.

The paragraph has been changed to: ***We calculated the number of matches using 1 to 14 pyramid levels and the scale factors 1.1, 1.2, 1.3 and 1.4. As a compromise between performance,***

i.e. number of matches, and computational efficiency (linked to the number of pyramid levels), a scale factor of 1.2 with seven pyramid levels was chosen. As shown in Figure 4, the number of matches does not increase significantly when using more than seven pyramid levels and even decreases towards 14 pyramid levels.

*) Page 6947, line 5: 'seven pyramid'

Agree, has been changed.

*) Page 6947, line 9: 'four representative'

Agree, has been changed.

*) Page 6947, line 11: 'Figures'

Agree, has been changed.

*) Page 6947, line 12: 'significantly'

Agree, has been changed.

*) Page 6947, line 24: 'longitude'

Agree, has been changed.

*) Page 6948, line 8: 'sufficiently computationally efficient'; Still, it sounds a bit awkward, but I have not come up with a better formulation. What do you consider sufficiently efficient?

The sentence has been changed to: ***The OpenCV feature tracking algorithms ORB, SIFT and SURF in combination with the python-toolbox 'Nansat' are computationally efficient (total processing time on regular MacBook Pro: 2–4 minutes) and allow high resolution sea ice drift retrieval from datasets with large temporal and spatial extent.***

*) Page 6948, lines 9 to 10: 'The processing times shown . . . '

Agree, has been changed.

*) Page 6948, line 12: '36 % and 67 %'

Agree, has been changed.

*) Page 6948, line 13: Comma after 'SURF'

Agree, has been changed.

*) Page 6948, lines 21 to 24: This sentence is not related to validation, so it should be moved to another section.

The sentence has been moved to the discussion.

*) Page 6949, line 4: ’This proves’; I think you can be more confident about your conclusion here.

Agree, has been changed.

*) Page 6949, line 6: ’Proofing’ is a noun only. What about ’showing’ or ’demonstrating’ or similar?

The sentence has been removed.

*) Page 6949, line 9: ’independent of’

The sentence has been removed.

*) Page 6949, line 13: What does ’This’ refer to? I assume you want it to refer to the ’not evenly distributed’ sea ice drift fields, while ’This’ actually refers to ’performance’. On a different note: Have you really used the uneven distribution of vectors in a quality measure in your manuscript? Maybe I misunderstood your quality measure. Please clarify.

The sentence has been removed and the term ’quality measure’ is not used anymore.

*) Page 6949, line 21: ’at no cost’

Agree, has been changed.

*) Page 6949, line 24: Comma after ’algorithm’

Agree, has been changed.

*) Page 6949, line 25: Comma after ’data’

Agree, has been changed.

4.7 Figures

*) Figures 1, 2, and 3: ’the four test...’

Agree, has been changed.

*) Figures 1 and 3: The captions do not mention the combination HH + HV.

The following has been added to the captions: ***Mean values of the four image pairs are shown in black and the sum of the mean values in red.***

and: ***Mean values are shown in black and the sum of the mean values in red.***

*) Figure 2, caption: Please clarify what you mean by 'chosen maximum'. I think it should read 'chosen parameter', because the value indicated by the grey line does not seem like the maximum to me. Mathematically, there seems to be only one maximum, therefore 'chosen maximum' confuses me.

Agree, 'chosen maximum' has been replaced by 'chosen parameter'.

*) Figure 4: The use of colour for the ocean and for Greenland and Svalbard is unnecessary. This makes the colour scale hard to interpret, because the colours on the map become a mix between the background colour and the colourbar. I think the background (that is, the ocean) should just be white. In the caption: 'grid cell', and comma after '(c)'. The figure size should be maximized, or the figure should be divided. Panel (a) is hard to interpret in print - even more so when I printed it in black/white - because of its many features. In the PDF version, magnifying is unproblematic because you use high-quality vector graphics, which is great.

The colour for ocean and land has been removed. The figures have been enlarged by removing 'a)', 'b)', 'c)' and the colour bar labels. The distance between the rows have been minimised. The vector width has been reduced. We kept the figure as one to facilitate the algorithm inter-comparison.

The caption has been changed to: ***Sea ice drift derived from Sentinel-1 image pair 'Fram Strait' (Table 2) using both HH and HV channel and ORB (first column, 6920 vectors), SIFT (second column, 1585 vectors) and SURF (third column, 518 vectors) algorithm. The panels show: drift vectors (red, first row), number of vectors per grid cell N (green, second row) and root mean square distance D in km (red, third row).***

*) Figure 5: For the figure to be useful, there must be a velocity scale. I assume the data in Fig. 5 is the same as in Fig. 4(a). However, some vectors present in Fig. 4(a) are not present in Fig. 5. Is that correct? If so, how and why have these been removed? See my comments on colour regarding Fig. 4. The use of red and green together is not a good choice and should be avoided.

The colour for ocean and land has been removed, a displacement scale has been added and green vectors have been replaced by white vectors. The ORB vectors, that were not used for calculating the root mean square error E, have been added to show the data in the same way as in Figure 8.

4.8 References

*) Bay et al. (2006): This is actually a book chapter. See here for the full bibliographic information: doi: 10.100711744023_32.

The reference has been changed to: ***Bay, H., Tuytelaars, T., and Van Gool, L.: Surf: Speeded Up Robust Features, in European Conference on Computer Vision, ISBN 978-3-540-33833-8, 404–417, doi: 10.100711744023_32, 2006.***

*) Korosov et al. (2015): I think the conference was called 'World Ocean Science Congress'. I could not find this reference or any proceedings from that conference. Can you make the presentation/poster/article available in the appropriate Nansen Center Github projects or on the Nansen Center's website?

A link to the poster from WOSC has been added to the Nansat main WIKI-page: https://github.com/nansencenter/nansat/wiki

*) Kwok (1998): The original title is all capitalized. Editors of the book are C. Tsatsoulis and R. Kwok.

The reference has been removed. See specific comments from referee #1.

*) Kwok (2010): 'sea-ice'

The reference has been removed. See specific comments from referee #1.

*) Kwok and Sulsky (2010): The abbreviated (and also full) journal title is 'Oceanography'.

The reference has been removed. See specific comments from referee #1.

*) Low (2004): The abbreviated journal title is 'Int. J. Comput. Vision'.

Agree, has been changed.

*) Rosin (1999): The abbreviated journal title is 'Comput. Vis. Image Und.'

Agree, has been changed.

*) Rosten and Drummond (2006): Similar to comment on Bay et al. (2006). See doi: 10.100711744023_34 for details.

The reference has been changed to: ***Rosten, E. and Drummond, T.: Machine learning for high-speed corner detection, in European Conference on Computer Vision, ISBN 978-3-540-33833-8, 430–443, doi: 10.100711744023_34, 2006.***

*) Rublee et al. (2011): I do not know what 'Willow Garage' is, but the paper was published in the journal IEEE I. Conf. Comp. Vis., see here: doi: 10.1109ICCV.2011.6126544.

The reference has been changed to: *Rublee, E., Rabaud, V., Konolige, K., and Bradski, G.: ORB: an efficient alternative to SIFT or SURF, IEEE I. Conf. Comp. Vis. (ICCV), ISBN: 978-1-4577-1101-5, 2564–2571, doi: 10.1109ICCV.2011.6126544, 6–13 Nov, 2011.*

4.9 Tables

*) Table 2: I do not understand the speed filter value. The setting you recommend is not a velocity. Besides, you apply both a low-speed and high-speed filter. What do you mean by 'tested range (resolution)'? I assume you mean 'tested range or resolution'. If this is what you mean, please do not use parentheses. The recommended settings for HV and HH channel brightness are not maxima in Fig. 2, even if the figure caption says so (see my comment below on the caption of Figure 2).

'Speed filter' has been replaced by *Displacement filter*. The lower displacement filter of 2.5 km is only applied during the testing phase to reject matches over land, but is not recommended for further application.
'(resolution)' has been replaced by *(increment)*. The value in parentheses refers to the increment or step size used during testing.
'Maximum' has been removed in the caption of Figure 3 and a reference to Equation 2 has been added. 'HH limits' and 'HV limits' in Table 2 have been replaced by *[*σ^0_{min}*,* σ^0_{max}*] (HH)* and *[*σ^0_{min}*,* σ^0_{max}*] (HV)*.

*) Table 3: 'Create two Nansat...'; 'from two Nansat objects'
Agree, has been changed.

Interactive comment

I would like to make you aware of an article by Komarov and Barber (2014), which may be a useful reference. Komarov and Barber (2014) used feature tracking to obtain sea ice drift from RADARSAT-2 images, and they also discussed the different results they obtained from the HV and HH channel.

The results from Komarov and Barber (2014) have been utilised and comparison to our findings has been added to the discussion section. See also General Comment 1.3.

Please find attached the corrected manuscript with changes marked in blue and red.

Thanks again for your comments. We are looking forward to your reply!

Best regards,

S. Muckenhuber, A. Korosov and S. Sandven

Manuscript prepared for J. Name
with version 2015/09/17 7.94 Copernicus papers of the LaTeX class copernicus.cls.
Date: 4 March 2016

~~Sea~~ Open source feature tracking algorithm for sea ice drift retrieval from Sentinel-1 SAR imagery~~using open source feature tracking~~

S. Muckenhuber, A.A. Korosov, and S. Sandven

Nansen Environmental and Remote Sensing Center (NERSC), Thormøhlensgate 47, 5006 Bergen, Norway

Correspondence to: S. Muckenhuber (stefan.muckenhuber@nersc.no)

Abstract. A ~~computational~~ computationally efficient, open source feature tracking algorithm, called ORB, is adopted and tuned for sea ice drift retrieval from Sentinel-1 SAR images. The best suitable setting and parameter values have been found using four ~~representative~~ Sentinel-1 image pairs ~~. A new quality measure for feature tracking algorithms is introduced utilising the distribution of the resulting vector field~~representative of sea ice conditions between Greenland and Severnaya Zemlya during winter/spring. The performance of the algorithm is compared ~~with~~ to two other feature tracking algorithms (SIFT and SURF). Applied on a test image pair acquired over Fram Strait, the tuned ORB algorithm produces the highest number of vectors (6920, SIFT: 1585 and SURF: 518) while being ~~computational~~ computationally most efficient (66 s, SIFT: 182 s and SURF: 99 s using a ~~2,7~~2.7 GHz processor with 8 GB memory). For validation purpose, ~~350~~ 314 manually drawn vectors have been compared with the closest calculated vectors, and the resulting root mean square ~~distance is 609.9~~error of ice drift is 563 m~~(equivalent to 7.5pixel)~~. All test image pairs show a ~~significant~~ significantly better performance of the HV channel due to higher informativeness. On average, around ~~4~~ four times more vectors have been found using HV polarisation. All software requirements necessary for applying the presented feature tracking algorithm are open source to ensure a free and easy implementation.

1 Introduction

Sea ice motion is an essential variable to observe from ~~earth observation~~ remote sensing data, because it strongly influences the distribution of sea ice on different spatial and temporal scales. Ice drift causes advection of ice from one region to another and export of ice from the Arctic Ocean to the sub-Arctic seas. Antarctic sea ice is even more mobile and its strong seasonality is linked to the ice transport from high to low latitudes (IPCC, 2013). Furthermore, ice drift generates convergence and divergence zones that cause formation of ridges and leads. ~~Presently,~~ However, there is still a lack of extensive sea ice drift data ~~do not provide~~ sets with sufficient resolution to estimate convergence and divergence ~~fields~~ on a spatial scaling of ~~a few~~ less than 5 kilometres.

The ~~main region of interest for this work~~ regions of interest are the ice covered ~~oceans~~ seas between Greenland and Severnaya Zemlya, i.e. Greenland Sea, Barents Sea, Kara Sea and the adjacent part of the Arctic Ocean. ~~These oceans are~~ This area is characterised by a strong seasonal cycle of sea ice cover, a large variation of different ice classes (Multi Year Ice, First Year Ice, Marginal Ice Zone etc.) and a wide range of drift speeds (e.g. strong ice drift in Fram Strait).

With systematic acquisition of space-borne Synthetic Aperture Radar (SAR) data over sea ice areas, ~~? and ?~~ Kwok et al. (1990) have demonstrated that high resolution ice drift fields can be derived from SAR data. SAR is an active microwave radar which ~~can acquire~~ acquires data independent of solar illumination and weather condition. Sea ice motion fields of the Arctic Ocean with a grid spacing of 5 km have been produced on a weekly basis between 1997–2012 using Radarsat and ENVISAT (Environmental Satellite) SAR data and the geophysical processor system introduced by ~~?~~ Kwok et al. (1990) . Thomas et al. (2008) have used pattern recognition to calculate sea ice drift between successive ERS-1 (European Remote-sensing Satellite) SAR images with a resolution of 400 m. This work has been continued by Hollands and Dierking (2011) using ASAR data from ENVISAT. Komarov and Barber (2014) used a similar pattern matching technique to evaluate ice motion results from dual-polarisation Radarsat-2 images.

With the successful launch of Sentinel-1A in April 2014 and the planned launch of Sentinel-1B in early 2016, high resolution SAR data will be delivered for the first time with open and free access for all users and ~~a never before reached repeat frequency~~ unprecedented revisit time of less than ~~1~~ one day in the Arctic (ESA, 2012). This introduces a new ~~area in SAR earth observation,~~ era in SAR Earth observation. Sea ice drift data with medium resolution (10 km) is provided operationally via the Copernicus Marine Environment Monitoring Service (CMEMS, http://marine.copernicus.eu), but no sea ice drift algorithm using Sentinel-1 data has been published so far. The objective of this paper is to identify and develop the most efficient open source ~~method~~ algorithm for high resolution sea ice drift retrieval from Sentinel-1 data.

Our goal is to exploit recent improvements and developments in computer vision by adopting a state of the art feature tracking algorithm to derive sea ice drift ~~. The advantage of feature tracking to algorithms based on pattern recognition is that each drift vector is independent of~~ (i.e. vectors of sea ice displacement). Current pattern matching algorithms constrain the high resolution vectors with low resolution estimates for practical reasons. Using feature tracking, drift vectors can be derived independently from the surrounding motion, which leads to better performance e.g. along shear zones. For application on large data sets and for operational use, ~~a computational~~ we considered a computationally efficient algorithm, called ORB ~~(Oriented FAST and Rotated BRIEF) (Rublee et al., 2011) , has been considered, tuned and compared~~ (Rublee et al., 2011) , tuned it for sea ice drift retrieval from Sentinel-1 imagery and compared the results with other available feature tracking algorithms and existing sea ice drift products.

The software requirements necessary for deriving ice drift fields from Sentinel-1 data (python with openCV and the python toolbox Nansat) are all open source to ensure a free, user friendly and easy implementation.

The paper is organised as follows: Section 2 introduces the used Sentinel-1A data product. The ORB algorithm description and the used methods for tuning, comparison and validation are presented in Section 3. The recommended parameter set including the tuning, comparison and validation results are provided in Section 4. The discussion can be found in Sect. 5.

2 Data

The Sentinel-1 mission, an initiative of the European Union and operated by the European Space Agency (ESA), is composed of a constellation of two identical satellites sharing the same near-polar, sun-synchronous orbit: Sentinel-1A, launched in April 2014, and Sentinel-1B, planned launch in early 2016. Sentinel-1 carries a single C-band Synthetic Aperture Radar (SAR) instrument ~~operating~~ measuring radar backscatter at a centre frequency of 5.405 GHz and supporting dual polarisation (HH+HV, VV+VH). With both satellites operating, the constellation will have a ~~repeat frequency~~ revisit time of less than 1 day in the Arctic. Radar data are delivered to Copernicus services within an hour of acquisition with open and free access for all users (ESA, 2012).

The Sentinel-1 product used in this paper is called “Extra Wide Swath Mode Ground Range Detected with Medium Resolution”. These images cover an area of 400 km $\times$ 400 km with a ~~resolution~~ pixel spacing of 40 m $\times$ 40 m (resolution: 93 m range $\times$ 87 m azimuth; residual planimetric distortions: within 10 m (Schubert et al., 2014)) and provide both HH and HV polarisation. Four image pairs (Table 1) representative ~~for~~ of our region of interest have been chosen, covering a range of different sea ice conditions (pack ice, fast ice, leads, ridges, marginal ice zone, ice edge etc.) and ~~time spans~~ intervals between the acquisitions. We focused on winter/spring data, since our area of interest experiences the highest sea ice cover during this period.

3 Method

Sentinel-1 datasets were opened and processed with the open source software Nansat (~~https://github.com/nansencenter/nansat)(Korosov et al., 2015)~~ see Appendix A; Korosov et al. (2015, 2016)). Nansat is a ~~scientist friendly~~ scientist-friendly Python toolbox for processing 2-D satellite ~~earth~~ Earth observation data. It is based on the Geospatial Data Abstraction Library (GDAL) and provides easy access to geospatial data, a simple and generic interface to common operations including reading, geographic transformation and export. Nansat proves to be efficient both for development and testing of scientific algorithms and for fast operational processing. To extend the functionality of GDAL, Nansat reads metadata from XML files accompanying Sentinel-1 data and supplements the GDAL data model with georeference information stored as ground control points (GCPs). Originally GCPs are pairs of latitude/longitude and corresponding pixel/line coordinates. In order to increase the accuracy of the geographic transformation, the projection of GCPs is changed from cylindrical to stereographic ~~centered at the center~~ centred at the centre of the scene. The reprojected GCPs are then ~~utilised~~ used by GDAL to calculate geographic coordinates of any pixel in the raster using spline interpolation. Reprojection of GCPs does not require much additional computational effort, but improves the result significantly, particularly ~~in~~ at high latitudes.

4 ~~Method~~

The normalised radar cross section (σ^0) is calculated from raw Sentinel 1A data using the following equation:

$$\sigma^0 = DN_i^2 / A_i^2 \qquad (1)$$

where DN_i is the digital number provided in the source TIFF file, A_i is the value of normalisation coefficient from the accompanying calibration metadata and i is an index of a pixel (Anonymous, 2014) . No additional pre-processing of SAR data was performed.

Our algorithm for sea ice drift detection includes three ~~major~~ main steps: (a) resampling of raw data to lower resolution, (b) detection and matching of features and (c) ~~a new introduced quality measure~~comparison/validation.

(a) To decrease the influence of speckle noise and increase the computational efficiency, the resolution is reduced before applying the ice drift algorithm ~~. Various resolution resampling algorithms have been tested: Gaussian, nearest neighbour, bilinear, cubic, cubic spline and Lanczos. Best performance and computational efficiency was achieved by using simple averaging~~ from 40 m to 80 m ~~resolution~~pixel spacing using simple averaging.

(b) For detection and tracking of features on large data sets and for operational use, a ~~computational~~ computationally efficient algorithm, called ORB ~~(Oriented FAST and Rotated BRIEF)~~(Rublee et al., 2011) , has been ~~utilised (Rublee et al., 2011)~~ used. In our numerical experiments we tuned the parameters of ORB for optimal SAR sea ice drift application. The best suitable parameter set (including spatial resolution of SAR image, patch size of FAST descriptor, number of pyramid levels, scale factor, etc.) has been evaluated ~~.~~ for our area and season of interest.

(c) ~~A new quality measure using the amount and deviation of vectors in a grid cell is introduced for feature tracking algorithms.~~ The introduced ORB setup is ~~tested against~~ compared to other available OpenCV feature tracking algorithms~~for comparison~~, CMEMS data and manually drawn vectors for performance appraisal and validation.

3.1 ORB algorithm

ORB (Oriented FAST and Rotated BRIEF) is a feature tracking algorithm introduced by Rublee et al. (2011) as 'a computationally-efficient replacement to SIFT ~~(Lowe, 2004) with~~ that has similar matching performance~~and~~, is less affected by image noise, and is capable of being used for real-time performance'. ORB builds on the FAST keypoint detector (Rosten and Drummond, 2006) and the binary BRIEF descriptor (Calonder et al., 2010) with many modifications to enhance the performance. It uses FAST to find ~~multiscale-features~~ multiscale-keypoints on several pyramid levels and applies a Harris corner measure (Harris and Stephens, 1988) to ~~find the best ones among them~~pick the best keypoints. To achieve rotation invariance, the orientation of the ~~feature~~ keypoint is calculated by using the ~~intensity weighted~~ intensity-weighted centroid of a circular patch with the located keypoint at the centre. ~~The ORB descriptor performs as well as SIFT and better~~ Rublee et al. (2011) states that the ORB descriptor performance is equal to SIFT (Lowe, 2004) and higher than SURF (Bay et al., 2006)~~, while being almost two orders of magnitude faster (Rublee et al., 2011) . To match features~~. Like Rublee et al. (2011) , we use ~~a Brute-Force matcher and Hamming-distance. An additional benefit of ORB is that it is free from licensing restrictions, unlike~~ Brute Force matcher and Hamming distance for feature matching. Unlike SIFT and SURF, ORB is an open-source software and use and distribution are not limited by any licenses.

Before the feature tracking algorithm can be applied ~~on~~ to a satellite image, the SAR backscatter values σ^0 have to be transformed into the intensity i range ~~0,255~~$(0 \leq i \leq 255 \text{ for } i \in \mathbb{R})$ used in openCV. This transformation is done by using Eq. (2) and setting all intensity values below and above the range to 0 and 255.

$$i = 255 \cdot \cancel{\frac{\sigma - \sigma_{\min}}{\sigma_{\max} - \sigma_{\min}}} \cdot \frac{\sigma^0 - \sigma^0_{\min}}{\sigma^0_{\max} - \sigma^0_{\min}} \tag{2}$$

Lower and upper brightness boundaries $\sigma^0_{\min}$ and $\sigma^0_{\max}$ are user defined and chosen to be constant in order to limit the influence of speckle noise and be independent of e.g. high backscatter values σ^0 over land. Converting the linear backscatter values ~~into decibel units~~ before the transformation into decibel units has been tested, but decreased the algorithm performance for both ~~HH and HV channel~~channels.

After the transformation into intensity values, keypoints are detected on both SAR scenes using the FAST-9 keypoint detector (Rosten and Drummond, 2006). FAST-9 compares the intensity ~~of a centre~~ I_p of a center pixel to the intensities of pixels on the surrounding circle with a perimeter of 16 pixels ~~. If 9 contiguous pixel~~ (Figure 1). If there exists a set of nine contiguous pixels in the circle ~~have an intensity difference greater (and with the same sign) than a certain threshold, the centre pixel is recoginsed as~~ which are all brighter than $I_p + t$, or all darker than $I_p - t$, the center pixel is recognized as a keypoint. The threshold t is set low enough to get more than the predefined amount ~~of retained~~ N of keypoints.

~~Each keypoint is addressed a score R using the intensity variation around the keypoint (Harris corner measure Harris and Stephens, 1988). A high intensity variation in both dimensions returns a high R value. The predefined amount of keypoints with the highest R values are utilised.~~

To detect features of different scales, the keypoint search is performed on several pyramid levels. The number of pyramid levels in combination with the scale factor defines the range and increment of the keypoint detection scaling. A scale factor of 2 means that each next pyramid level has 4 times less pixels, but such a large scale factor degrades the feature matching score. On the other hand, a small scale factor close to 1 means to cover a certain scale range needs more pyramid levels and hence, the computational cost increases.

~~ORB adds an orientation θ~~ FAST does not produce a measure of cornerness and Rublee et al. (2011) have found that it has large responses along edges. Harris corner measure (Harris and Stephens, 1988) is used to order the FAST keypoints according to their cornerness and reject less reliable keypoints. Considering a window $w(x,y)$ around the keypoint, the intensity derivatives I_x, I_y in x and y direction can be written in a matrix $\mathbf{M}$:

$$\mathbf{M} = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \tag{3}$$

The eigenvalues λ_1 and λ_2 of $\mathbf{M}$ contain the intensity derivative in the direction of the fastest and slowest change, respectively. Based on λ_1 and λ_2, a score R can be calculated for each keypoint:

$$R = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \tag{4}$$

with k being an empirical constant. A high intensity variation in both dimensions returns a high R value. The top N keypoints with the highest R values are used and the rest is rejected.

FAST does not include orientation, but ORB adds a direction to each keypoint ~~, derived from connecting the keypoint and the intensity weighted centroid of a~~ using the intensity-weighted centroid from Rosin (1999) . Using the moments m_{pq} of a circular area around the keypoint,

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y) \tag{5}$$

the intensity-weighted centroid has its location at

$$C = \left(\frac{m_{01}}{m_{00}}, \frac{m_{10}}{m_{00}}\right) \tag{6}$$

The orientation θ (e.g. green arrow in Figure 1) represents the direction of the vector connecting the keypoint with the intensity-weighted centroid. The moments m_{pq} are computed with x and y

remaining within a circular region of radius r, where r is chosen to be the size of the patch p ~~with the keypoint at the centre (Rosin, 1999)~~ used for the following feature description Rublee et al. (2011) .

~~To describe keypoints,~~ After locating and adding orientation to the best N keypoints, a patch p around each keypoint is used for feature description (NB: keypoint refers to 1 pixel, feature refers to description of p). ORB applies a modified version of the binary ~~keypoint~~ descriptor BRIEF (Calonder et al., 2010). ~~A~~ Rublee et al. (2011) defines a binary test τ ~~on~~ for a patch p ~~is defined by:~~ as follows:

$$\tau(p;\cancel{x}X,\cancel{y}Y) := \begin{cases} 1 & \text{if } p(\cancel{x}X) < p(\cancel{y}Y) \\ 0 & \text{if } p(\cancel{x}X) \geq p(\cancel{y}Y), \end{cases} \tag{7}$$

~~where $p(x)$ is the intensity value at point x. A feature $\boldsymbol{f}$ can be described by a vector of~~ with $p(X)$ and $p(Y)$ being the intensities at test point X and Y, respectively. ORB uses 5x5 sub-windows as test points (e.g. in Figure 1). Applying n binary tests ~~:~~ on a single patch, Rublee et al. (2011) derive a binary feature vector $\boldsymbol{f}$:

$$\boldsymbol{f}_n(p) := \sum_{1 \leq i \leq n} 2^{i-1} \tau(p;\cancel{x}X_i,\cancel{y}Y_i) \tag{8}$$

~~A~~ The considered set of n binary tests with ~~sampling pair location (x_i, y_i)~~ test points (X_i, Y_i) can be written in a 2 x n matrix ~~:~~ (Rublee et al., 2011) :

$$\mathbf{S} = \begin{pmatrix} X_1, \ldots, X_n \\ Y_1, \ldots, Y_n. \end{pmatrix} \tag{9}$$

To be invariant to in-plane rotation, Rublee et al. (2011) steers $\mathbf{S}$ according to the orientation θ using the corresponding rotation matrix $\mathbf{R}_\theta$:

$$\mathbf{S}_\theta = \mathbf{R}_\theta \mathbf{S}. \tag{10}$$

A good set $\mathbf{S}$ of sampling pairs needs to be uncorrelated, so that each pair adds new information to the descriptor, and have high variance, to make features more discriminative. Rublee et al. (2011) applied a greedy search on a large training dataset to obtain a set ~~of 256~~ for ORB with $n = 256$ relatively uncorrelated tests with high variance.

After ~~detection and description,~~ the feature description, openCV allows different matching procedures for ORB. Like Rublee et al. (2011) , we use Brute Force matching and compare each feature of the first image ~~is compared~~ to all features in the second image~~(Brute Force matching) using the~~ .

As a comparison measure, we use the Hamming distance which is equal to the number of positions in which the two considered feature vectors have a different value~~(Hamming distance).~~ . E.g. comparing the two binary vectors $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$

$$\boldsymbol{b_1} = 1011101$$
$$\boldsymbol{b_2} = 1001001 \tag{11}$$

returns the Hamming distance $d = 2$, since the third and fifth position have a different value.

Our setting returns the best two matches and applies the ratio test from Lowe (2004) to decide whether the best match is accepted or rejected. The match is accepted if ratio of the distances $\frac{d_1}{d_2} <$ is below a given threshold. The ratio test eliminates a high number of false matches, while discarding only few correct matches.

3.2 ORB setting and parameter tuning

Achieving the best possible performance of ORB for sea ice drift from Sentinel-1 images, requires a good setting and tuning of the parameters shown in Table 2.

~~As a compromise between performance and computational efficiency, the resolution of the Sentinel-1 image is reduced from 40 to 80m (resize factor=0.5) and the amount of maximum retained keypoints is set to 100000.~~

It is not recommended to re-project one image onto the projection of the second image before applying the ORB algorithm, since this is ~~computational~~ computationally very expensive. Instead, geographic coordinates of the matched start and end point shall be calculated independently using the georeference information from GCPs of the first and second image.

~~To reject less reliable matches, we use the ratio test explained in Lowe (2004) .~~ Manual interpretation of ice drift results (using the training data from Table 1) ~~, suggest~~ reveals that a good compromise between amount of vectors and correct results ~~with a~~ can be achieved with a Lowe ratio test threshold ~~of~~ equal to 0.75~~, meaning that the Hamming-distance~~. That means that the Hamming distance of the best match has to be less than 0.75 $\times$ ~~Hamming-distance~~ Hamming distance of the second best match. Tested on the image pairs from Table 1, the ratio test showed clear better performance and is ~~computational~~ computationally less expensive than the alternative cross-check, where ~~keypoints~~ features are matched in both directions (first image to second image and vice versa) and rejected ~~, if the matches are not correlated~~ if the drift vectors are too different.

~~Unreasonable high velocities above 10~~ Unreasonably high sea ice displacements (e.g. above 40 km ~~+1kmh^{-1}~~ for a time difference between two scenes of ~30 hours) are removed in a post-processing step ~~of~~ from the drift field. In addition, displacements below 2.5 km are rejected during the testing to disregard matches over land. This does not influence the number of correct matches, since the sea ice displacement in all considered test images is above 2.5 km.

~~To tune the remaining parameters in Table 2,~~ Based on our observations we assume that the ~~amount~~ proportion of wrong matches ~~relative to the amount of correct matches~~ does not increase with increasing total number of matches. ~~Using this assumption , it can be concluded that more matches equals better algorithm performance~~ Under this assumption the algorithm performance refers to the total number of matches and is used to tune the algorithm parameters in Table 2. ORB is ~~computational very efficient , making it possible to test the remaining~~ computationally more efficient enabling testing the parameters over a wide range with high resolution using both HH and HV polarisation. ~~This has been done to find the best suitable values for patch size, HH and HV brightness boundaries, pyramid levelsand scale factor.~~

3.3 ~~Quality measure~~

As a starting point, the tested parameters were set as follows: resize factor = 0.5, patch size = 31, pyramid levels = 8, scale factor = 1.2, HH limits = [0,0.12], HV limits = [0,0.012] and ratio test = 0.8. As a compromise between performance and computational efficiency, the amount of maximum retained keypoints is set to 100 000. Tested range and parameter meaning are shown in Table 2.

~~The resulting sea ice drift vector field is not evenly distributed, but according to the recognition performance of~~ In order to find an optimal value for the tested parameter, it is varied in reasonable a range, the ~~respective area. Regions with few vectors represent low reliability, whereas regions with many vectors suggest high reliability. By using a grid and calculating the amount~~ feature tracking algorithm is applied and the total number of matched vectors is found. Once the best suitable value for a tested parameter is found, it is applied for further testing.

3.3 Comparison of ORB to SIFT and SURF

The presented ORB algorithm has been compared to other OpenCV feature tracking algorithms, namely SIFT (Lowe, 2004) and SURF (Bay et al., 2006) , using image pair “Fram Strait” (Table 1). SIFT and SURF were used in standard mode and the framework conditions were set equal for the comparison. Image pre-processing has been done as described above, Brute Force Matching including the Lowe ratio test with threshold 0.75 has been applied for all three algorithms as well as the removal of unreasonably high sea ice displacements in a post-processing step. Since SIFT allows to define the number of retained keypoints, this parameter has been set to 100 000 as done for ORB. The further tuning of SIFT and SURF is not the aim of this paper, since these two algorithms are not open source and computationally less efficient.

The distribution and reliability of the calculated vector fields have been assessed using two parameters on a grid with cell size 1 °longitude × 0.2 °latitude: number of derived vectors per grid cell (N) and root mean square distance (~~RMSD~~D) of all vectors ~~appearing in one~~ in a gird cell computed as follows:

$$D = \sqrt{\frac{\sum_i (u_i - \tilde{u})^2 + (v_i - \tilde{v})^2}{N}} \tag{12}$$

where i is the index of a vector inside the grid cell, ~~the distribution of the vector field can be used as a quality measure~~u_i and v_i are the eastward and northward drift components and $\tilde{u}$, $\tilde{v}$ the corresponding mean values.

4 ~~Results~~

3.1 Validation

~~Before testing the individual parameters, the remaining parameters were set to the following values:pyramid levels~~The ORB algorithm has been validated against drift data from two independent sources using the image pair “Fram Strait” (Table 1). First, 350 features have been identified by a sea ice expert in both images and manually connected using ArcGIS. Second, sea ice drift vectors were taken from the Copernicus Marine Environment Monitoring Service (CMEMS, http://marine.copernicus.eu). The SAR ice drift product of CMEMS is operated by the Technical University of Denmark (DTU) and drift data is provided with a resolution of 10 ~~=~~km using pattern matching techniques (Pederson et al. (2015) , http://www.seaice.dk/).

Since the starting location of ORB, manual and CMEMS vectors do not coincide, the corresponding (ORB) reference vectors were found as nearest neighbours within 5 ~~8, scale factor=1.2, HH limits=, HV limits=, ratio test=0.8. An additional low speed filter with 2.5km is applied during the testing to reject matches over land. This filter does not influence the number of correct matches, since the sea ice velocities in all considered test images are above 2.5km. Once the best suitable value for a certain parameter is found, it is applied for further testing.~~ km radius from the (CMEMS or manual) validation vectors.

Three parameters were considered for the comparison: root mean square error (E), slope (S) and offset (O) of the linear fit between the reference and validation vectors. E was calculated as follows:

$$E = \sqrt{\frac{\sum_i (u_i - U_i)^2 + (v_i - V_i)^2}{n}} \tag{13}$$

where i is the index of a vector pair (reference and validation vector) inside the entire sample, u_i and v_i are eastward and northward drift components of the validation vector, U_i and V_i are eastward and northward components of the reference vector and n is the number of vector pairs.

In addition, the CMEMS data has been validated against manual vectors in order to understand the credibility of the reference data.

4 Results

4.1 ORB parameter tuning

Table 2 shows the recommended parameter set for ORB Sentinel-1 sea ice drift application for our region and period of interest. Using these parameters yielded the best compromise between performance and computational efficiency for the four representative image pairs from Table 1.

4.1.1 Patch size

Figure 2 shows that ~~patch size values~~ changing the size (length and width) of the considered patch p between 10 and 60 ~~pixel can vary~~ pixels can modify the resulting amount of vectors by an order of magnitude. To resolve ~~velocity~~ drift gradients with high resolution, the patch size shall be as small as possible. Taking ~~that~~ this into account and the performance represented by the amount of matches, the best suitable patch size was chosen to be 34 ~~pixel (2.72~~pixels . For our training dataset (Table 1), this yields on average around 1 and 4 vectors per 10 ~~km). This yields around 1000 and 6000 vectors~~ km^2 for HH and HV~~respectively for the training dataset from Table 1~~, respectively. The four image pairs respond similar to a patch size variation. 'Franz Josef Land' has the highest number of HH matches and the lowest for HV.

4.1.2 Brightness boundaries

The performance of the algorithm (represented by the amount of matches) for different backscatter limits ~~σ_{max} and σ_{min}~~ σ^0_{max} (Equation 2) for HH and HV polarisation is shown in Fig. 3. Within the chosen backscatter range, the amount of vectors can vary by an order of magnitude. As a compromise between the different results of the four image pairs, ~~the maximum backscatter σ_{max} is suggested to be set~~ we suggest to set the upper brightness boundary σ^0_{max} to 0.08 and 0.013 for HH and HV. The chosen ~~minimum σ_{min}~~ lower boundary σ^0_{min} is 0 for both HH and HV, ~~since~~ because the number of matches ~~is decreasing towards higher values for most training images~~decreases for increasing values of σ^0_{min} (not shown). Applying this setting on the training dataset yields on average around ~~1500 and 6000 vectors~~ 1 and 4 vectors per 10 km^2 for HH and HV.

4.1.3 Pyramid levels and scale factor

~~Figure 4 displays~~ We calculated the number of matches using ~~1–16~~ 1 to 14 pyramid levels and the scale factors 1.1, 1.2, 1.3 and 1.4. As a compromise between performance, i.e. number of matches, and computational efficiency (linked to the number of pyramid levels), a scale factor of 1.2 with ~~7~~ seven pyramid levels was chosen.

4.1.4 ~~Recommended parameter set~~

~~Table 2 shows the recommended parameter set for ORB Sentinel-1 sea ice drift application. Using these parameters yielded the best compromise between performance and computational efficiency for the 4 representative image pairs from Table 1~~As shown in Figure 4, the number of matches does not increase significantly when using more than seven pyramid levels and even decreases towards 14 pyramid levels.

4.2 HH and HV comparison

~~Figure~~Figures 2, 3 and 4 display the HH and HV results with solid and dashed lines, respectively. All image pairs show ~~a significant~~ significantly better performance of the HV channel. On average, around 4 times more vectors have been found using HV. Even the image pair “Franz Josef Land” (Table 1), which has the best HH and the worst HV performance, shows more than two times more vectors using HV channel. However, due to the different appearance of sea ice in the HH and HV image, the spatial distribution of the resulting drift vectors is also slightly different~~, supporting the usefulness of a combination of both results.~~.

4.3 ~~Comparison with SIFT and SURF~~

~~To compare the introduced ORB setup with other available OpenCV feature tracking algorithms, SIFT (Lowe, 2004) and SURF (Bay et al., 2006) are considered. The performance of the three algorithms is tested on the~~ Figure 6 shows the spatial distribution of identified keypoints and matched features in a 200×200 pixels sub-image from image pair “Fram Strait” (Table 1)~~using the introduced quality indexes number of vectors (N) and root mean square distance (RMSD)on a grid with cell size 1°Longitude~~. The results for HH and HV are displayed in two separate panels. The density of identified keypoints in HH (11 keypoints per 10×10 pixels window) is in the same order of magnitude as in HV (15 keypoints per 10×10 pixels window). This is expected, since the number of retained keypoints for both channels is set to 100 ~~0.2° latitude.~~

~~Figure 8a shows the combined vector fields of the HH and HV channel for ORB, SIFT and SURF respectively. Figure 8b and c display N and RMSD on the considered grid.~~ 000 for the entire scene. However, the number of matched features in HH is significantly lower (0.15 features per 10×10 pixels window) than in HV (1.6 features per 10×10 pixels window). The observed difference in matching success can be explained by looking at the frequency distribution of the radar backscatter standard deviation in a sliding window with same size as used for feature description (34×34 pixels). The comparison in Figure 7 shows that HH provides a few windows with very high variability, i.e. high standard deviation, but the majority has very low backscatter variability (sharp peak with mode 20). On the HV image however, most of the windows have a medium to high backscatter variability (wide peak with mode 25) which is more favourable for keypoint detection.

4.3 Comparison with SIFT and SURF

A total of 6920, 1585 and 518 vectors are found for the image pair “Fram Strait” (Table 1) using ORB, SIFT and SURF~~respectively.~~, respectively (Figure 8 a). Comparing the vector fields using N and D, as described in Section 3, shows that ORB covers the largest area with ~~many (here >50)~~ more vectors per grid cell ~~and corresponding low RMSD values~~ (on average $N > 50$, Figure 8 b) and lower corresponding root mean square distance values (on average $D < 1.5$ km, Figure 8 c).

Comparing the distributions of N (Q-Q plot in Figure 9, left panel), shows that ORB derives in all cases around 5 to 10 times more vectors than SIFT and SURF. The Q-Q plot in the right panel of Figure 9 considers the distributions of D. For $D < 400\,m$, the vectors derived by ORB exhibit a higher variability within one grid cell (slightly higher D), probably due to a larger number of vectors N. For the higher root mean square values ($D > 400\,m$), SIFT and SURF vectors are much less consistent than ORB vectors (higher D).

4.4 Computational efficiency

The OpenCV feature tracking algorithms ORB, SIFT and SURF in combination with the python-toolbox 'Nansat' are ~~sufficient computational efficient to compute~~ computationally efficient (total processing time on regular MacBook Pro: 2–4 minutes) and allow high resolution sea ice drift ~~fields~~ retrieval from datasets with large temporal and spatial extent. The processing ~~time indications~~ times shown in Table 3 are based on testing the algorithms on a MacBook Pro from early 2013 with a 2.7 GHz Intel Core i7 processor and 8 GB 1600 MHz DDR3 memory. Applying the introduced ORB algorithm needs 36 % and 67 % of the processing time necessary to compute drift fields with SIFT and SURF, respectively.

4.5 Validation

~~In order to validate the calculated ORB drift field shown in Fig. 8~~Since reference vectors were searched only within a given radius of the validation vectors, the number of match-ups decreased for the ORB vs manual comparison from 350 possible matches to 314, for ORB vs CMEMS from 560 to 436 and for CMEMS vs manual from 350 to 201 (Table 4). The average distance between compared vectors was 1702 m, 2261 m and 3440 m for ORB vs manual, ORB vs CMEMS and CMEMS vs manual, respectively.

The validation of ORB vectors with manual derived vectors (Figure 5 a, ~~350 features have been identified by a sea ice expert in both images and manually connected using ArcGIS.~~ Table 4) reveals a high accuracy of our tuned ORB algorithm with root mean square error $E = 563$ m, slope $S = 1.02$ and offset $O = -372$ m. Given the displacement range for the used image pair of 10–35 km, the relative error of the algorithm (ratio of E to mean drift) is 2.5%.

The vector distributions of ORB and CMEMS (Figure 5~~shows the manually drawn vectors (green) and the respective nearest neighbour vectors from ORB (red) . The resulting RMSD between manual and calculated vectors is 609.9 m (equivalent to 7.5~~b) are similar. ORB covers a larger area in total, but in a few regions only CMEMS provides drift information. The ORB vs CMEMS comparison gives an error $E = 1641$ ~~pixel)~~ m, slope $S = 1.03$ and offset $O = 265$ m (Table 4).

Validating CMEMS using manual data results in the highest root mean square error $E = 1690$ m with slope $S = 0.98$ and offset $O = -415$ m (Table 4) . ~~The RMSD represents~~

Decreasing the threshold radius between reference and validation vectors does not influence the error E significantly but reduces the number of found matching vectors especially when comparing CMEMS and manual vectors.

5 Discussion and outlook

The open source feature tracking algorithm ORB (Oriented FAST and Rotated BRIEF) has been tuned for sea ice drift retrieval from Sentinel-1 SAR imagery and used for processing winter/spring data in the ice covered oceans between Greenland and Severnaya Zemlya. Validating calculated drift results against manual derived vectors, we found that our algorithm ($E_{ORB} = 563\,m$) had a distinctly higher accuracy than the drift dataset provided by CMEMS ($E_{CMEMS} = 1690\,m$). The given root mean square errors E represent a combination of ~~the manually produced error and the displacement variation between the manual and calculated vector . Using the~~ three error sources:

- error of manual ice drift identification introduced by the sea ice expert;
- difference between derived and reference vector due to different geographical location of the starting point (maximum 5 km);
- actual error of the algorithm.

Hence, the actual error of the tuned ORB algorithm ~~, a total of~~ is expected to be even lower than 563 m.

As expected, the application of the tuned ORB algorithm is much more efficient than manual ice drift assessment: e.g. 6920 vectors have been calculated within 3 minutes, whereas identifying 350 sea ice drift vectors manually takes several hours. The number of calculated vectors can be increased by returning a higher number of keypoints (e.g. 1 000 000). However, the processing time increases proportional to the square of the considered keypoints and the algorithm performance becomes suboptimal at some point.

The presented ORB algorithm ~~outperforms~~ also outperforms other available feature tracking algorithms, such as SIFT and SURF not only in processing time, but also in quantity and quality of drift vectors, measured by the two introduced indexes ~~N and RMSD. This shall proof~~ N and D. This proves that ORB is the best option for feature tracking of sea ice on Sentinel-1 SAR imagery. ~~In addition, the ORB results compare very well with manually drawn drift vectors, proofing good reliability of the algorithm.~~

~~Current algorithmsfor calculating~~ The algorithm tuning has been performed using winter/spring data, since our area of interest experiences the highest sea ice cover during this period. During summer/autumn, most considered areas have very little or no ice cover (e.g. Barents Sea and Kara Sea), making ice drift calculation during this period less meaningful. Nevertheless, some areas, like the western Fram Strait, experience sea ice cover during the entire year. Dependence of the algorithm performance on the season needs to be evaluated in future work. Computing sea ice drift from summer/autumn data is expected to be more demanding, since features might be destroyed by melting.

Comparing the four considered image pairs, 'Franz Josef Land' yields the highest number of HH matches, accompanied by the lowest number from HV channel. A distinct shorter time difference between the acquisitions (8 hours for 'Franz Josef Land' compared to more than 30 for the other image pairs) might be one reason for an improved HH performance. That would conclude that HH features are less preserved over time and increasing the repeat frequency of the satellite (as planned with Sentinel-1B) will improve the algorithm performance in particular for the HH channel. The sea ice conditions are another important factor, when comparing the algorithm performance for different scenes. The image pair 'Fram Strait' includes the marginal ice zone in the eastern part and multi year ice in the north west. Not many matches are expected in the marginal ice zone, but the multi year ice includes more stable deformation pattern, like ridges, that lead to a good feature tracking performance. 'Svalbard North' includes a very small part of the marginal ice zone and the major part is comparable homogeneous pack ice with long cracks along a prevailing direction. 'Franz Josef Land' and 'Kara Sea' are clearly less homogeneous and show a mixture of ice floes with different scales and newly formed young ice. This paper has been focusing on finding the best suitable algorithm for a range of ice conditions found in the considered area and we can give an idea how ice conditions and acquisition time might affect the ORB feature tracking performance. Further investigations needs to be done in order to evaluate the algorithm performance for different ice conditions and other areas like the Beaufort Sea or Antarctica.

Komarov and Barber (2014) have evaluated sea ice drift results from dual-polarisation Radarsat-2 imagery using a combination of phase and cross-correlation. Comparing the polarisation channels, HH is more sensitive to small-scale roughness, whereas the HV channel provides more stable, large scale features linked to ice topography. Komarov and Barber (2014) concludes that the combination of HH and HV is beneficial, since more reliable vectors are provided and the vector distributions complement each other. They also found that noise floor stripes in the HV images do not affect the motion tracking from pattern matching. We can extent this discussion for feature based algorithms. Using noise removal for HV and angular correction for HH has been tested, but did not improve the feature tracking results, i.e. a lower number of vectors has been found. Like Komarov and Barber (2014) , we recommend the usage of both channels since the vector distributions are complementary. However, using feature tracking, HV provides about 4 times more vectors than HH, making HV the more informative channel. The different performance can be explained by a higher variability of the HV backscatter intensity, considering a window with the same size as used for feature description (34×34 pix).

Contemporary algorithms for calculating sea ice drift vectors from consecutive image pairs are based either on feature tracking or pattern ~~recognition. Feature tracking provides vectors, which are independent from each other, whereas pattern recognition includes the surrounding drift information . Therefore pattern recognition is more prone to errors in areas with high velocity gradients. The resulting drift fields from feature tracking are generally not evenly distributed, but according to~~

~~the feature recognition performance of the respective area.This can be utilised for quality measure (as shown in this paper), but is a disadvantage when it comes to estimation of divergence, shear and deformation. Pattern recognition algorithms however,~~ matching. The feature tracking approach detects keypoints on two images based solely on the backscatter distribution of the images without taking other keypoints into account. Hence, ORB identifies the keypoints independently. Based on the keypoint locations, the binary feature vectors are calculated. During the second step, all features in the first images are compared to all features in the second image without taking drift information from surrounding vectors into account, i.e. the matching of features from one image to the other is also done independently. Although very close keypoints may share some pixels during the feature description process (i.e. overlap of the considered patches around the keypoints), the detection of keypoints and matching of features are done independently. Eventually, feature tracking vectors are independent of each other in terms of position, lengths and direction, allowing very close drift vectors to point into different directions.

Figure 10 illustrates 430 drift vector anomalies detected in a 300×400 pix (24×32 km) sub-image from “Fram Strait” (Table 1) close to the marginal ice zone. The anomalies are calculated as difference to the mean drift of the entire scene. This example shows that very small scale dynamic processes, such as the observed rotation, can be detected and quantified with the feature tracking approach.

Common pattern matching techniques limit the independence of neighbor vectors for practical reasons. First, pattern matching is usually performed on a regular grid, determining the position and distance between vectors. Second, pattern matching often follows a pyramid approach in order to speed up processing (Thomas et al., 2008) : low resolution drift is initially estimated using large sub-windows and large steps. This first guess constrains the following pattern matching at finer scale. Repeating this procedure increases the resolution of the end product, but length and direction of the high resolution vectors depend on the low resolution estimates, i.e. neighbor vectors depend on each other. Although pattern matching can be designed to retrieve independent vectors by varying the extent of the correlation area and the spacing between vectors, for practical reasons the overlap between the correlation areas is usually half the size of the area (Thomas et al., 2008b) .

The independence of feature tracking vectors has positive and negative implications. On one hand, very close vectors, that are independent in length and direction, allow identification of ice deformation at very high resolution. The variogram (Fig. 11), which shows how vector differences dependent on the distance between them (Cressie, 1993) , indicates, that very close vectors may differ significantly, although the difference is generally linearly proportional to the distance. On the other hand, feature tracking vectors are not evenly distributed in space, and large gaps may occur between clouds of densely located vectors. Spatial irregularity is not optimal for systematic detection of divergence and shear zones and calculation of deformation.

Therefore, computationally efficient feature tracking should be complemented by systematic pattern matching to deliver evenly distributed, high resolution vector fields. ~~Hence, combining~~ Combining the two different drift calculation approaches and making use of the respective advantages is ~~recommended to improve high resolution ice drift estimation~~planned as the next step of our research.

Appendix A: Open source distribution

The presented work is entirely based on open source software (Python, openCV and Nansat) and satellite images with open and free access for all users. Sentinel-1 SAR data can be downloaded ~~for~~ at no cost, in near real time under https://scihub.esa.int/dhus/. The used programming language is Python, a free and open source software available under https://www.python.org. The OpenCV (Open Source Computer Vision) programming library includes the ORB algorithm, and a python compatible version can be downloaded under http://opencv.org. To handle and read the satellite data, Nansat is used, which is a scientist friendly Python toolbox for processing 2-D satellite Earth observation data (source code incl. installation description can be found under https://github.com/nansencenter/nansat). The presented sea ice drift algorithm including an application example can be downloaded from https://github.com/nansencenter/sea_ice_drift.

Acknowledgements. This research was supported by the Norwegian Research Council project IceMotion (High resolution sea-ice motion from Synthetical Aperture Radar using pattern tracking and Doppler shift, project number 239998/F50). We thank both developer groups of the ORB feature tracking algorithm and the ~~earth~~ Earth observation toolbox Nansat. The ~~utilised~~ used satellite data was provided by the European Space Agency. A special thanks to Wolfgang Dierking (AWI) for valuable input that improved the paper.

References

Anonymous: https://earth.esa.int/web/sentinel/technical-guides/sentinel-1-sar, 2014

Bay, H., Tuytelaars, T., and Van Gool, L.: Surf: Speeded ~~up robust features,~~ Up Robust Features, in European Conference on Computer Vision, ~~LNCS 3951~~ISBN 978-3-540-33833-8, 404–417, ~~Graz, Austria,~~ doi: 10.100711744023 SUBSCRIPTNB32, 2006.

Calonder, M., Lepetit, V., Strecha, C., and Fua, P.: BRIEF: Binary Robust Independent Elementary Features, CVLab, EPFL, Lausanne, Switzerland, 2010.

Cressie, N.: Statistics for spatial data: Wiley series in probability and statistics, Wiley-Interscience New York, 15, 16, 1993.

ESA: Sentinel-1 ESA's Radar Observatory Mission for GMES Operational Services, ESA Communications, SP-1322/1, ISBN: 978-92-9221-418-0, ISSN: 0379-6566, 2012.

Harris, C. and Stephens, M.: A Combined Corner and Edge Detector, Plessey Research Roke Manor, UK, The Plessey Company, 147–152, 1988.

Hollands, T. and Dierking, W.: Performance of a multiscale correlation algorithm for the estimation of sea-ice drift from SAR images: initial results, Ann. Glaciol., 52, 311–317, 2011.

IPCC – Intergovernmental Panel on Climate Change: Climate Change 2013: The Physical Science Basis, Fifth Assessment Report, AR5, 317–382, 323–335, 2013.

Komarov, A.S., and Barber, D.G.: Sea Ice Motion Tracking From Sequential Dual-Polarization RADARSAT-2 Images, IEEE Transactions on Geoscience and Remote Sensing, Vol. 52(1), No. 1, 121–136, doi: 10.1109/TGRS.2012.2236845, 2014.

Korosov, A. A., Hansen, M. W., and Yamakava, A.: Nansat – scientist friendly toolbox for processing satellite data, World Ocean ~~Scientific Congress,~~ Science Congress, Korosov, A. A., Hansen, M. W., and Yamakava, A. Cochin, India, 2–8 February, 2015.

~~Kwok, R.: The RADARSAT geophysical processor system, Analysis of SAR data of the polar oceans: recent advances, Berlin, Springer-Verlag, 235–257, 1998.~~

Korosov, A.A. Hansen, W.M., Yamakawa, A., Dagestad, K.F., Vines A., Riechert, M., Myasoedov, A., Morozov, E.A., Zakhvatkina, N.: Nansat-0.6.8, Zenodo, doi:10.5281/zenodo.45703, 2016.

~~Kwok, R.: Satellite remote sensing of sea ice thickness and kinematics: a review, J.Glaciol., 1129–1140, 2010.~~

Kwok, R.~~and Sulsky, D.: Arctic ocean Sea Ice Thickness and Kinematics: Satellite Retrievals and modeling, J. Oceanogr. Soc., 23, 134–143~~, ~~2010.~~ Curlander J.C., McConnell R., and Pang S.: An Ice Motion Tracking System at the Alaska SAR Facility, IEEE Journal of Oceanic Engineering, Vol. 15, No. 1, 44–54, 1990.

Lowe, D. G.: Distinctive image features from scale-invariant keypoints, Int. J. ~~Comp~~Comput. Vision, 60, 91–110, 2004.

Pedersen, L.T., Saldo, R. and Fenger-Nielsen, R.: Sentinel-1 results: Sea ice operational monitoring, Geoscience and Remote Sensing Symposium (IGARSS), IEEE International, 2828–2831, doi=10.1109/IGARSS.2015.7326403, 2015

Rosin, P. L.: Measuring corner properties, ~~Comp. Vision Image Underst~~Comput. Vis. Image Und., 73, 291–307, 1999.

Table 1. Sentinel-1 image pairs used for parameter tuning~~, date shown in year month day – hour minute second~~.

Region	~~Acquisition date 1st image~~ 1-st image acquisition time, UTC	~~Acquisition date 2nd image~~ 2-nd image acquisition time, UT
Fram Strait	~~2015/03/28 –~~ 2015-03-28 07:44:33	~~2015/03/29 –~~ 2015-03-29 16:34:52
Svalbard North	~~2015/04/22 –~~ 2015-04-22 06:46:23	~~2015/04/23 –~~ 2015-04-23 13:59:03
Franz Josef Land	~~2015/~~2015-03-24 03~~/24 – 03~~:21:13	~~2015/03/24 –~~ 2015-03-24 11:30:06
Kara Sea	~~2015/04/22 –~~ 2015-04-22 11:37:16	~~2015/04/24 –~~ 2015-04-24 11:20:59

Table 2. Recommended ~~parameter~~ set of parameters for ~~ORB Sentinel-1~~ retrieval of sea ice drift ~~application~~from Sentinel-1 data using ORB.

Parameter	Meaning	Tested range (increment)	Recommended setting
Amount keypoints	Maximum number of keypoints to retain	–	100 000
Resize factor	Resolution reduction during pre-processing	0.5–1 (0.5)	0.5
Patch size	Size of descriptor patch in pixels	10–60 (1)	34
Pyramid levels	Number of pyramid levels	1–15 (1)	7
Scale factor	Pyramid decimation ratio	1.1–1.4 (0.1)	1.2
$[\sigma^0_{\mathrm{min}}, \sigma^0_{\mathrm{max}}]$ (HH)	Brightness boundaries for HH channel	[0–0.04,0.01–0.2] (0.01)	[0,0.08]
$[\sigma^0_{\mathrm{min}}, \sigma^0_{\mathrm{max}}]$ (HV)	Brightness boundaries for HV channel	[0–0.007,0.001–0.02] (0.001)	[0,0.013]
Ratio test	Threshold for ratio test	0.5–1 (0.1), 0.7–0.8 (0.01)	0.75

Rosten, E. and Drummond, T.: Machine learning for high-speed corner detection, in European Conference on Computer Vision, ~~LNCS 3951~~ISBN 978-3-540-33833-8, 430–443, ~~Graz, Austria,~~ doi: 10.100711744023 SUBSCRIPTNB34, 2006.

Rublee, E., Rabaud, V., Konolige, K., and Bradski, G.: ORB: an efficient alternative to SIFT or SURF, ~~Willow Garage, Menlo Park, California,~~ IEEE I. Conf. Comp. Vis. (ICCV), ISBN: 978-1-4577-1101-5, 2564–2571, doi: 10.1109ICCV.2011.6126544, 6–13 Nov, 2011.

Schubert, A., Small, D., Meier, E., Miranda, N., and Geudtner, D.: Spaceborne Sar Product Geolocation Accuracy: A Sentinel-1 Update, Geoscience and Remote Sensing Symposium (IGARSS), IEEE International, 2675–2678, doi=10.1109/IGARSS.2014.6947025, 2014

Thomas, M., Geiger, C. A., and Kambhamettu, C.: High resolution (400 m) motion characterization of sea ice using ERS-1 SAR imagery, Cold Reg. Sci. Technol., 52, 207–223, 2008.

Thomas, M., Geiger, C.A., Kannan, P., Kambhamettu, C.: Streamline Regularization for Large Discontinuous Motion of Sea Ice, in 2008 IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS 2008), ed. by Selim Aksoy and Nicolas H. Younan, International Association for Pattern Recognition, Institute of Electrical and Electronics Engineers, Inc., Tampa, FL, 1–4, doi: 10.1109/PRRS.2008.4783171, 2008.

Table 3. Processing ~~time of steps during computation of~~ times for sea ice drift ~~field~~ computation from ~~Sentinel-1 imagery, NB: per~~ one channel~~(apart from creating Nansat object)~~.

Process	Time [s]
Create ~~2~~ two Nansat objects from Sentinel-1 image pair	21.1
Read matrixes from ~~2~~ two Nansat objects	48.8
Apply feature tracking algorithm – ORB	65.8
Apply feature tracking algorithm – SIFT	181.8
Apply feature tracking algorithm – SURF	98.5

Table 4. Comparison of ORB, CMEMS and manual derived sea ice drift data from image pair 'Fram Strait' (Table 1). The total numbers of derived vectors are 6920 (ORB), 560 (CMEMS/DTU) and 350 (manual). The # vector pairs is the number of used vector pairs for comparison, i.e. vector pairs with maximum 5 km distance. The average distance refers to the starting locations of the used reference and validation vectors. E is the root mean square error, S and O are slope and offset of the linear fit.

Algorithm	E [m]	S	O [m]	# vector pairs	Average distance [m]
ORB vs manual	563	1.02	-372	314	1702±1325
ORB vs CMEMS	1641	1.03	265	436	2261±1247
CMEMS vs manual	1690	0.98	-415	201	3440±1105

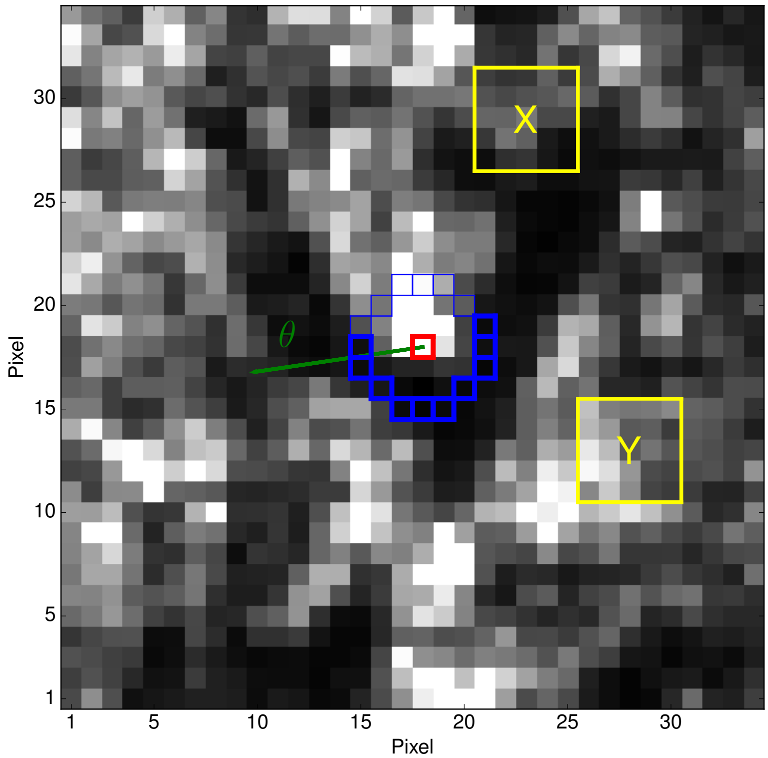

Figure 1. Subset of the 1^{st} image from 'Fram Strait' pair (Table 1) with centre at 2.31°W, 81.70°N and pixel spacing of 80 m. The centre pixel (red) is recognised as keypoint since $\geqslant 9$ contiguous pixels (bold blue) of the surrounding blue circle have intensity values smaller than the centre minus threshold t. The orientation θ of the keypoint is shown with a green arrow. The displayed area (34x34 pixels) around the keypoint represents the considered patch p used for feature description. The yellow 5x5 pixels sub-windows X and Y are an example for a possible binary test sampling pair with $p(X) < p(Y)$ and hence, $\tau(p; X, Y) = 1$ (Equation 7).

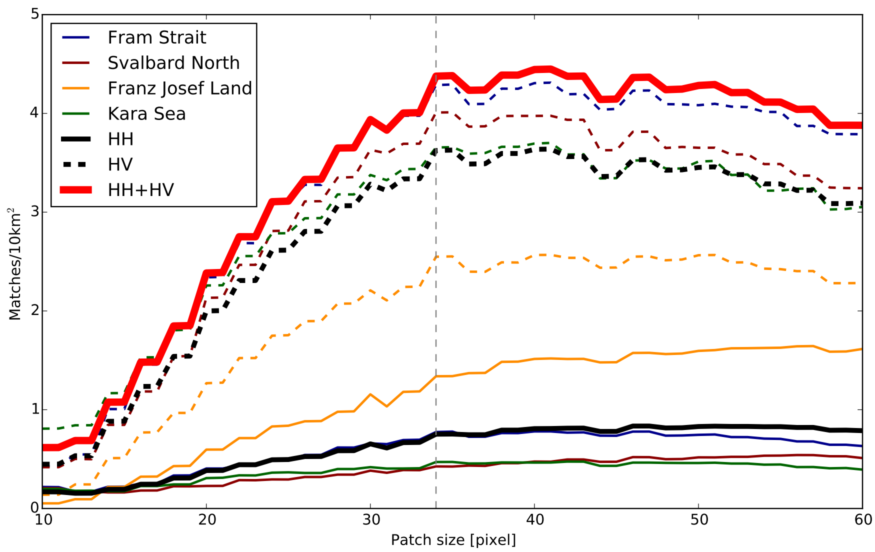

Figure 2. Patch size of descriptor versus ~~average~~ number of matches of the ~~4~~ four test image pairs from Table 1. Solid and dashed lines represent results for HH and HV polarisation, respectively. Mean values of the four image pairs are shown in black and the sum of the mean values in red. Vertical grey line at 34 ~~pixel (2.72km)~~ pixels represents chosen parameter.

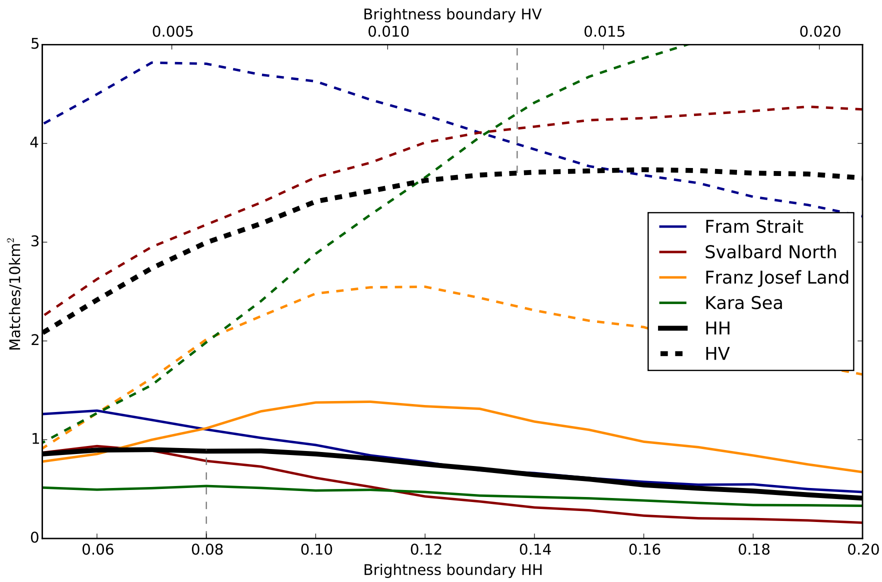

Figure 3. ~~Maximum~~ Upper brightness boundary σ^0_{max} (Equation 2) versus ~~average~~ number of matches of the ~~4~~ four test image pairs from Table 1. Solid and dashed lines represent results for HH and HV, respectively. Black lines are the mean values of the four image pairs. Vertical grey ~~line~~ lines at 0.08 (HH) and 0.013 (HV) ~~represents~~ represent chosen ~~maximum~~parameters.

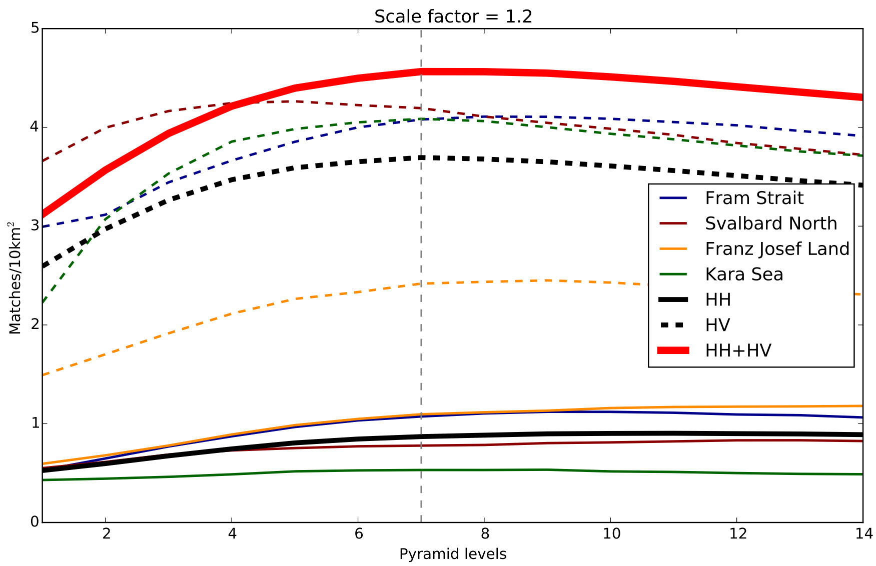

Figure 4. Number of pyramid levels versus ~~average~~ number of matches of the ~~4~~ four test image pairs from Table 1 for a scale factor ~~ofb~~ of 1.2. Solid and dashed lines represent results for HH and HV polarisation. Mean values are shown in black and the sum of the mean values in red. Vertical grey line at 7 represents chosen number of pyramid levels.

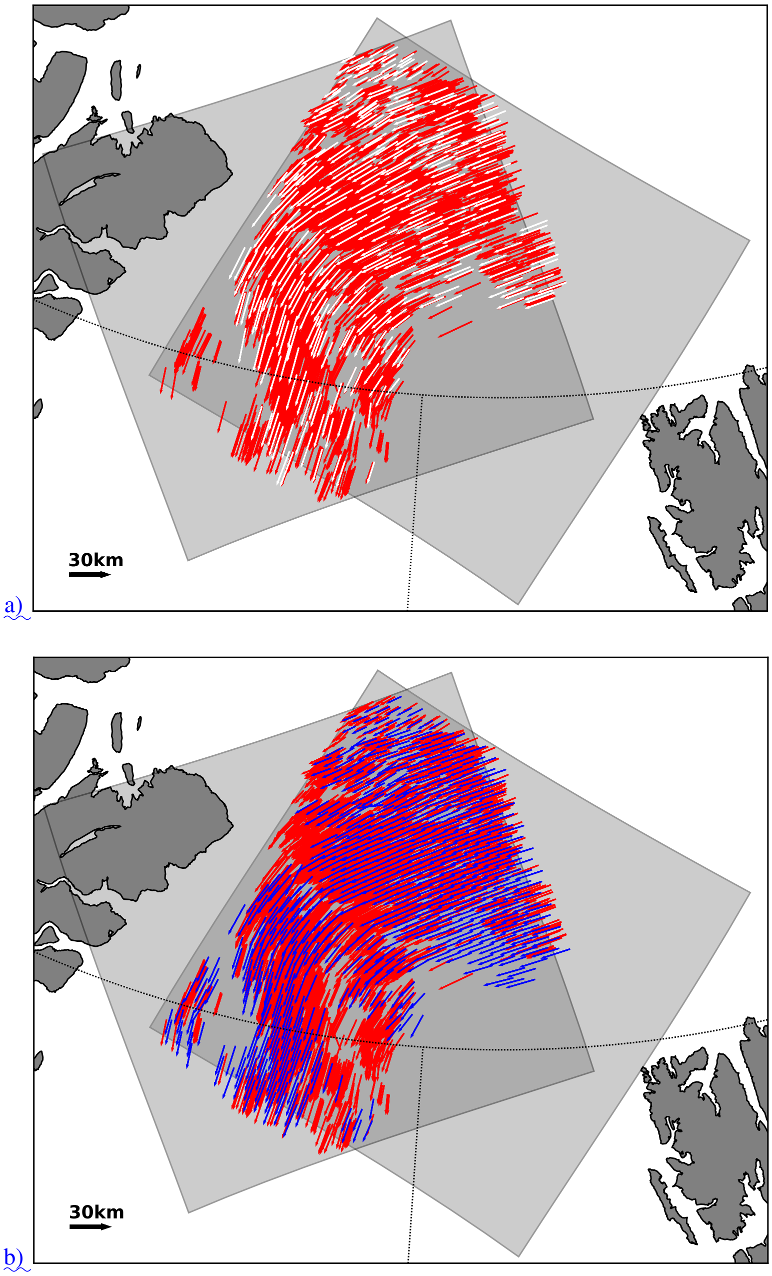

Figure 5. Sea ice drift of the Sentinel-1 image pair “Fram Strait” (Table 1). (**a**) Manually drawn vectors are shown in ~~green~~ white and the ~~respective nearest neighbour~~ computed ORB vectors in red. (**b**) shows ORB vectors in comparison to the drift vectors from the CMEMS/DTU data (blue).

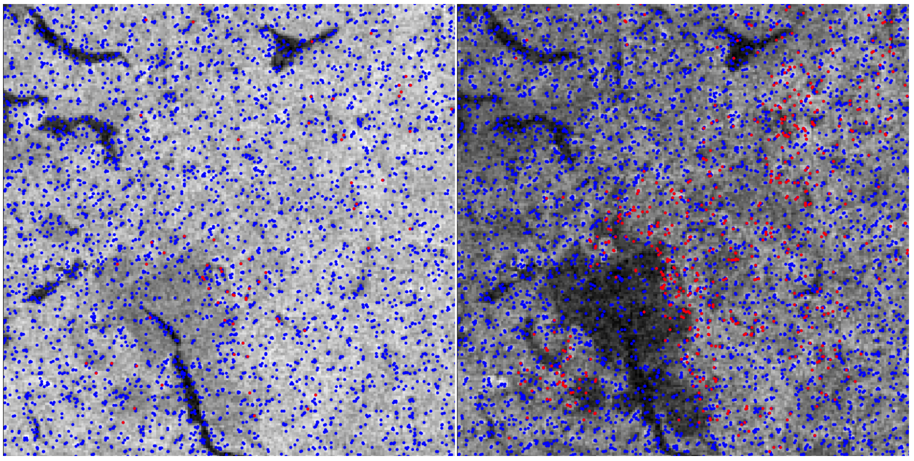

Figure 6. Identified keypoints (blue) and matched features (red) on a 200×200 pixels sub-image from the pack ice area in image pair “Fram Strait” (Table 1). Results of HH are shown in the left panel and HV in the right panel.

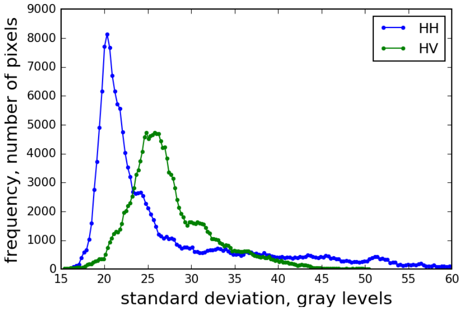

Figure 7. Frequency distribution of radar backscatter standard deviation using a 34×34 pixels sliding window (step = 1 pixel) on a 1000×1000 pixels sub-image from image pair “Fram Strait” (Table 1). The radar backscatter is scaled to range 0–255 using Equation 2. The considered sub-image covers pack ice, marginal ice zone and small parts of open water. Results for HH are shown in blue and HV in green.

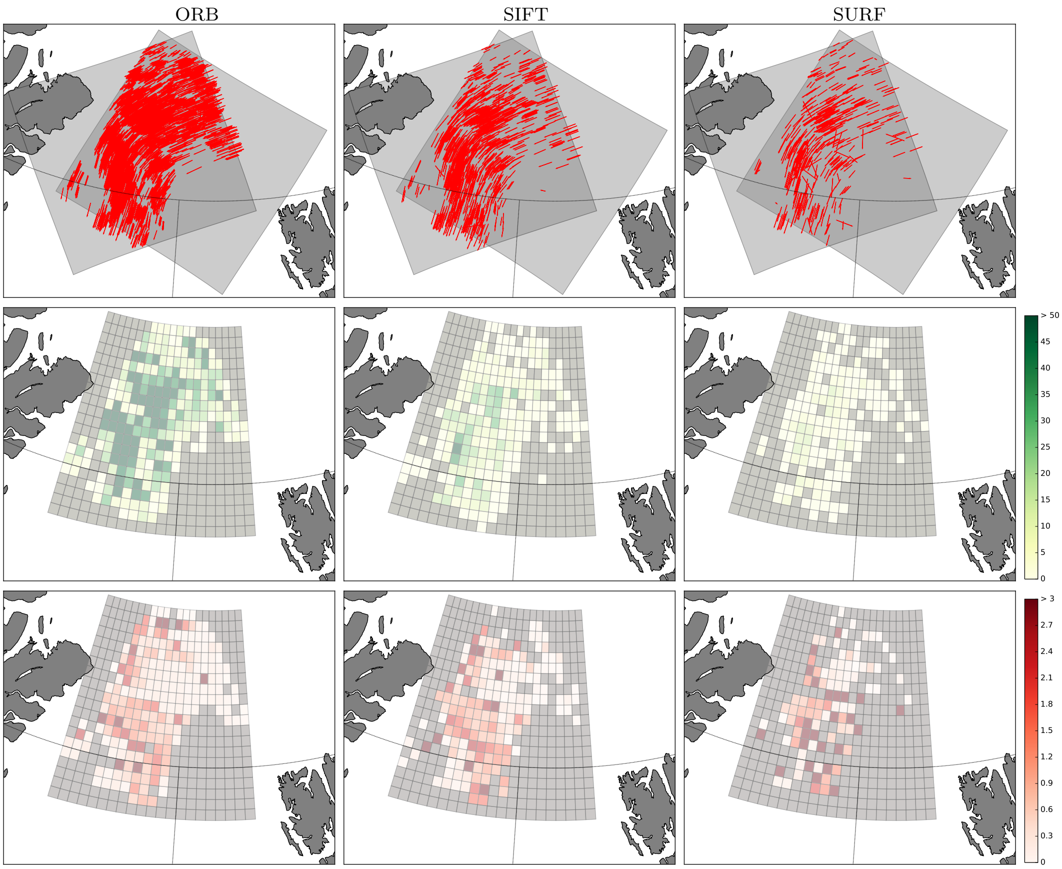

Figure 8. Sea ice drift derived from Sentinel-1 image pair 'Fram Strait' (Table 1) using both HH and HV channel and ORB (first column, 6920 vectors), SIFT (second column, 1585 vectors) and SURF (third column, 518 vectors) algorithm. The panels show: drift vectors (red, first row), number of vectors per grid cell N (green, second row) and root mean square distance D in km (red, third row).

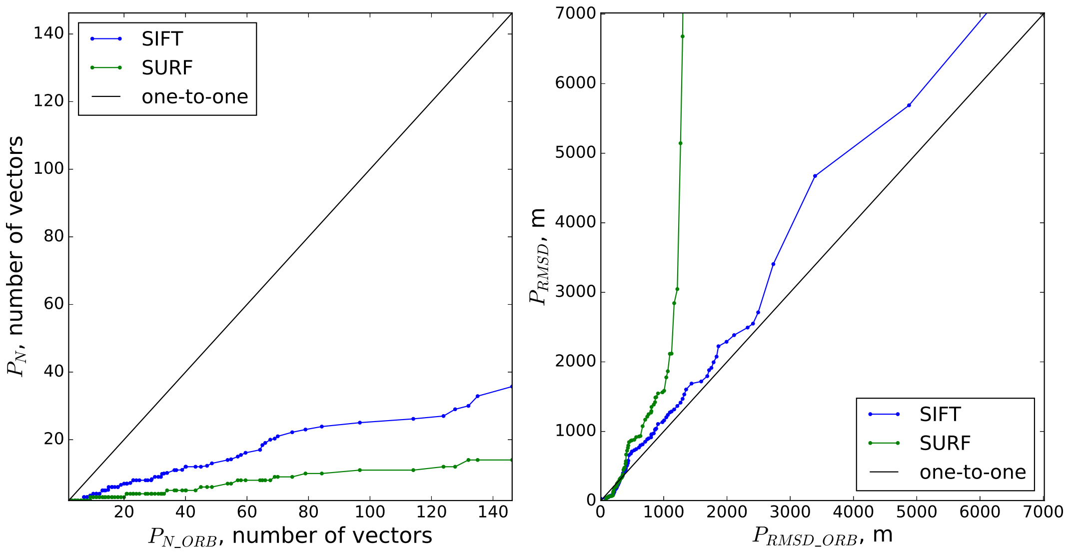

Figure 9. Q-Q plot of number of vectors N (left panel) and root mean square distance D (right panel) from results shown in Figure 8. Tuned ORB algorithm (X-axis) compared to SIFT (Y-axis, blue dots) and SURF (Y-axis, green dots).

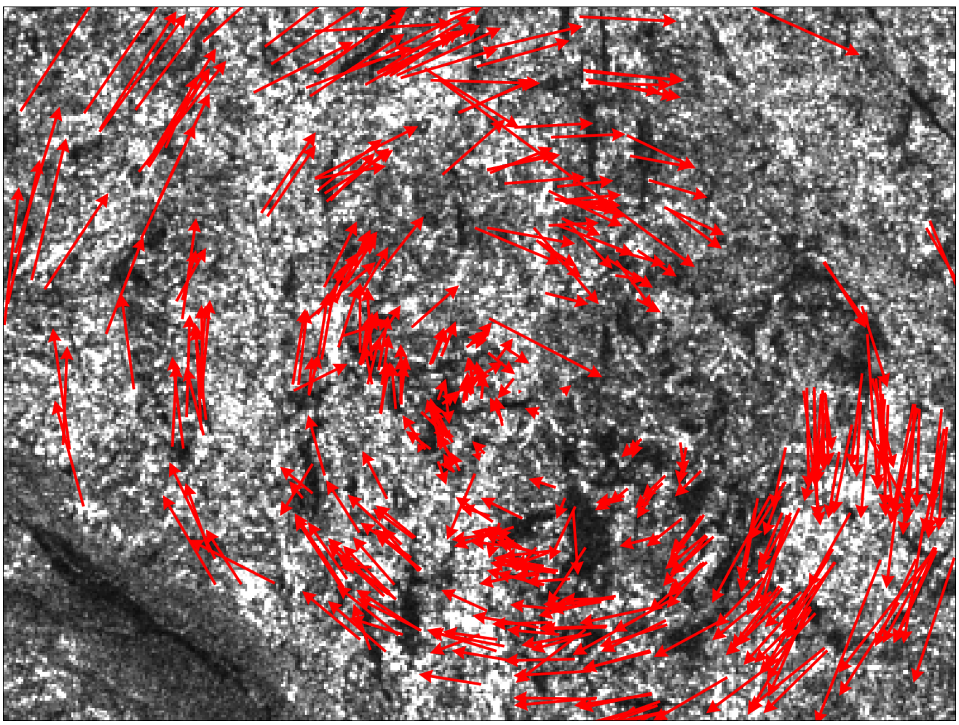

Figure 10. Sea ice drift anomaly (compared to mean drift of the scene) detected in a 300×400 pix (24×32 km) sub-image from "Fram Strait" (Table 1) close to the marginal ice zone.

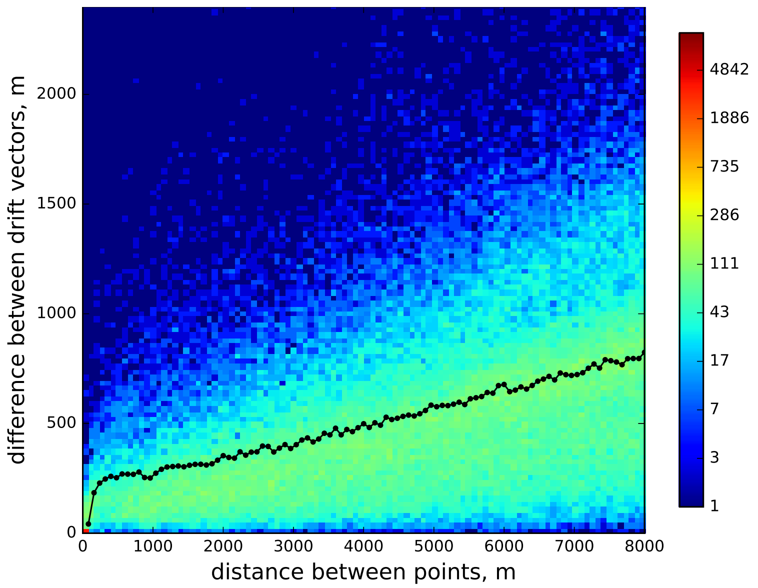

Figure 11. Variogram of drift vectors (black line) on top of 2D histogram of distance between vectors and difference between vectors estimated from vectors identified on the Sentinel-1 image pair "Fram Strait" (Table 1). Colour of the 2D histogram indicates the number of vectors.