Development of a machine learning technique for automatic analysis of seafloor image data: Case example, Pogonophora coverage at mud volcanoes

A. Lüdtke [a,*], K. Jerosch [b,c], O. Herzog [a], M. Schlüter [c]

[a] Center for Computing and Communication Technologies (TZI), Universität Bremen, Am Fallturm 1, D-28359 Bremen, Germany
[b] Bedford Institute of Oceanography, 1 Challenger Drive (P.O. Box 1006), Dartmouth, NS, Canada B2Y 4A2
[c] Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, D-27570 Bremerhaven, Germany

ABSTRACT

Digital image processing provides powerful tools for fast and precise analysis of large image data sets in marine and geoscientific applications. Because of the increasing volume of georeferenced image and video data acquired by underwater platforms such as remotely operated vehicles, means of automatic analysis of the acquired image data are required. A new and fast-developing application is the combination of video imagery and mosaicking techniques for seafloor habitat mapping. In this article we introduce an approach to fully automatic detection and quantification of Pogonophora coverage in seafloor video mosaics from mud volcanoes. The automatic recognition is based on textural image features extracted from the raw image data and classification using machine learning techniques. Classification rates of up to 98.86% were achieved on the training data. The approach was extensively validated on a data set of more than 4000 seafloor video mosaics from the Håkon Mosby Mud Volcano.

© 2011 Published by Elsevier Ltd.

1. Introduction

Taking, analyzing, and processing digital images have become important tasks for marine scientists to cope with photographs, satellite images, and geographic information systems (GIS) techniques, caused by a significant increase in image data describing the seafloor. Seabed mapping has been on the rise since the widespread implementation of multibeam sonar, providing marine scientists with georeferenced three-dimensional depictions of seafloor morphology and sediment properties. Multibeam sonar combined with high-resolution digital video information obtained from underwater vehicles such as remotely operated vehicles (ROVs), autonomous underwater vehicles (AUVs), and submersibles enable the creation of benthic habitat maps. These maps are essential to enhance knowledge about and protection of the ocean, as well as its sustainable management. Compared with bathymetric measurements, high-resolution video images provide much more detailed and small-scale information about the seafloor. Also, bathymetric multibeam sonar surveys can cover much larger areas than video surveys, which require a low speed of the research vessel. Therefore, the spatial coverage of video material is unavoidably more fragmentary than bathymetric maps. Compared to the amount of valuable image and video material, only a small portion—often only defined sequences—is analyzed manually in a time-consuming way.

The increasing application of ROVs and AUVs for scientific and engineering objectives results in multimedia information requiring cost-efficient and effective annotation and content-based retrieval methods. Thus, the development of automatic content-based analysis methods for georeferenced image data from the underwater domain came to the fore of marine science. This study focuses on further developments of the analysis framework that we introduced in Jerosch et al. (2007), involving the application of machine learning techniques for feature detection, combined with scientific visualization and analysis tools such as GIS, for recognition of characteristic seafloor features at submarine mud volcanoes. Submarine mud volcanoes are characterized by upward flow of mud from deeper geological strata, enhanced temperatures, high inventories and fluxes of methane and other hydrocarbons, and occurrence of specific chemoautotrophic communities (Hjelstuen et al., 1999; Kopf, 2002; Milkov, 2000).

The enhancements of the software target the detection not only of bacterial mats (Jerosch et al., 2007) but also of tube worms living on the seafloor at the Håkon Mosby Mud Volcano (HMMV), located at a depth of 1250 m in the center of the most recent Bear Island fan slide valley (see Fig. 1). Both bacterial mats (Beggiatoa) and tube worms (Pogonophora) are strongly related to the enzymatic oxidation of reduced compounds of microbial anaerobic methane oxidation (AOM; Boetius et al., 2000). Detecting and quantifying AOM areas is an essential task, as AOM is a major biological sink of methane in marine sediments. It is crucial in maintaining a sensitive balance of the greenhouse gas content of the Earth's atmosphere (Hinrichs and Boetius, 2002).
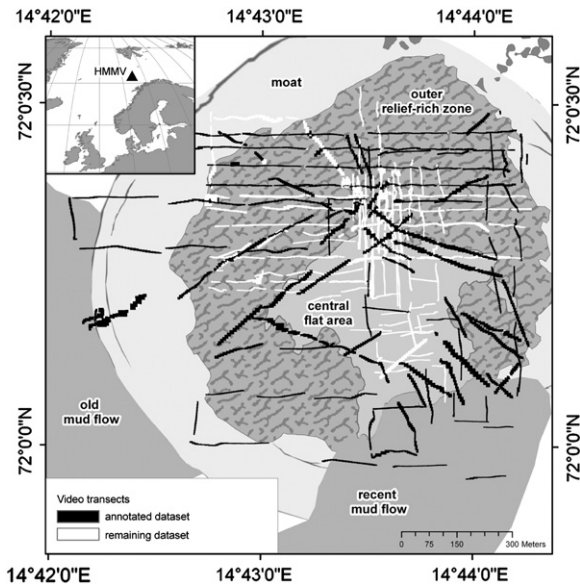
Fig. 1. The Håkon Mosby Mud Volcano is about 1.4 km in diameter at water depths of 1250–1266 m (Vogt et al., 1997). It is a concentric morphologic structure with highly gas-saturated sediments. A flat central area of gray fluid-like mud with a high geothermal gradient (Kaul et al., 2006) is surrounded by a zone of bacterial mats. This center is surrounded by elevated sediment features (relief-rich zone) densely populated by Pogonophora tube worms (Jerosch et al., 2006a). This inset shows location of the Håkon Mosby Mud Volcano northwest of Norway.

2. Related work

One of the major objectives in the field of automatic image analysis applied to marine sciences is to detect, identify, and track various marine species and seafloor characteristics in underwater video footage acquired by fixed or towed cameras. This includes the possibility of using these techniques for future work on ROVs and AUVs.

The analysis of still photographs and videos in marine sciences is mostly performed manually (e.g., Jerosch et al., 2006b; Kostylev et al., 2001). Automatic feature detection is not often applied in marine sciences (Guinan et al., 2009; Jerosch et al., 2007; Kohler and Gill, 2006; Zhanga et al., 2005) since the implementation of suitable algorithms is particularly complicated. If the video data come from a towed system, algorithms have to deal with sparse and unstable lighting due to the limitation of energy, and varying speed, angle, and altitude of the camera above the seafloor. The quality of a video stream coming from a ROV is enhanced by a more stable speed, more constant height over ground, and more homogeneous lighting conditions (no energy limitations). For both systems, the algorithms have to track a broad spectrum of known and unknown features occurring at the seafloor.

Due to the variety of technical conditions, quality, and resolution an automatic analysis either is not possible or is more laborious than its benefit justified. In the case of the HMMV, the development of an automatic analysis system was appropriate because of the large amount of acquired image data. Furthermore, the HMMV is one of the well-known submarine mud volcanoes of high latitudes and is the focus of marine scientists from all over the world.

The HMMV is a site of ongoing fundamental research regarding its geology (Feseker et al., 2008; Hjelstuen et al., 1999, 2007; Kaul et al., 2006; Milkov et al., 2004; Perez-Garcia et al., 2009; Vogt et al., 1997, 1999), marine geochemistry (Damm and Budéus, 2003; Lein et al., 1999; Sauter et al., 2007), (micro-)biology (de Beer et al., 2006; Gebruk et al., 2003; Lösekann et al., 2007, 2008; Niemann et al., 2006;

Omoregie et al., 2008; Pimenov et al., 1999; Purser et al., 2009; Smirnov, 2000; Soltwedel et al., 2005), and biogeochemical habitats (Jerosch et al., 2006a; Milkov et al., 1999). Several expeditions have been and will be undertaken considering the HMMV as a long-term case study.

3. Acquisition of the underlying image data set

w?>During six dives by the ROV Victor[1] (IFREMER, France), image data were recorded and processed using MATISSE (mosaicking advanced technologies integrated in a single software environment; Allais et al., 2004; Vincent et al., 2003; see Fig. 2). This software, developed by IFREMER, creates mosaics from a video stream provided by a camera mounted vertically on the ROV and accomplishes these images with navigation data supplied by acoustic USBL (ultrashort baseline) positioning (Vincent et al., 2003).

To ensure high-quality imagery, the surveys were performed at an altitude of 3 m above the seafloor. Since the aperture of the camera is 60°, the width of the mosaics is about 3 m. The image size varies due to the mosaicking process. On average, the scale is at 0.9162×0.9162 cm per pixel, varying slightly due to varying height over ground. The speed of the ROV was 0.3 m/s at maximum. A single mosaic consists of 500 merged images collected within half a minute during the mosaicking process. It covers an area of about $3 \times (6-7)$ m. As the mosaics are not rectangular in general, the output images contain regions where no data is available. Such regions are marked black (see Fig. 3b). Within a treatment (a row of consecutive images), successive images overlap in small parts of their data regions (see Fig. 4). The entire image data set consists of 4108 georeferenced video mosaics and covers $\approx 8\%$ of the HMMV area (see Fig. 1).

## 4. Learning visual properties of image regions with Pogonophora coverage

Recognition of Pogonophora seafloor coverage is achieved by supervised machine learning. Data items are represented by so-called features vectors, vectors of numerical image features extracted from the raw data in our case. A classification function using feature vectors as input and a class as output is inferred based on supervised (manually prepared) training data. Four commonly applied machine learning classifiers of different types have been evaluated on the data set (see also Table 1):

1. Support vector machines (Vapnik, 1995): A functional classifier based on the construction of discriminating hyperplanes in high-dimensional spaces to which the feature vectors are transformed.
2. K-nearest-neighbors classifier (Aha and Kibler, 1991): An instance-based learning scheme based on distances of instances to classify to training instances.
3. C4.5 decision trees (Quinlan, 1993): A "divide-and-conquer," tree-based statistical classifier.
4. Naive Bayes classifier (John and Langley, 1995): A probabilistic classifier based on Bayes's rule.

All implementations have been taken from the WEKA machine learning toolkit[2] (Witten and Frank, 2005). All of the above learning schemes have in common that—using different techniques—a partition of the feature space is computed for classification. While support vector machines are probably the most popular choice

---
[1] See http://www.ifremer.fr/fleet/systemes_sm/engins/victor.htm.
[2] See http://www.cs.waikato.ac.nz/ml/weka.
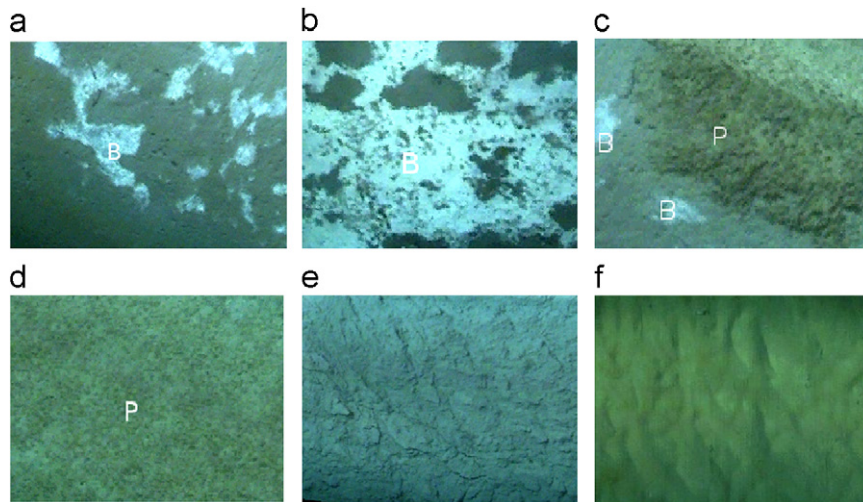
Fig. 2. Examples of different habitat types at HMMV (*Beggiatoa* and Pogonophora marked with labels "B" and "P"): (a) *Beggiatoa* patches 20–50%, (b) *Beggiatoa* mats > 50%, (c) *Beggiatoa* patches < 20% and Pogonophora 50–80%, (d) Pogonophora > 80%, (e) 100% structured mud, and (f) 100% smooth mud with ripples. Images recorded by Victor6000/IFREMER.
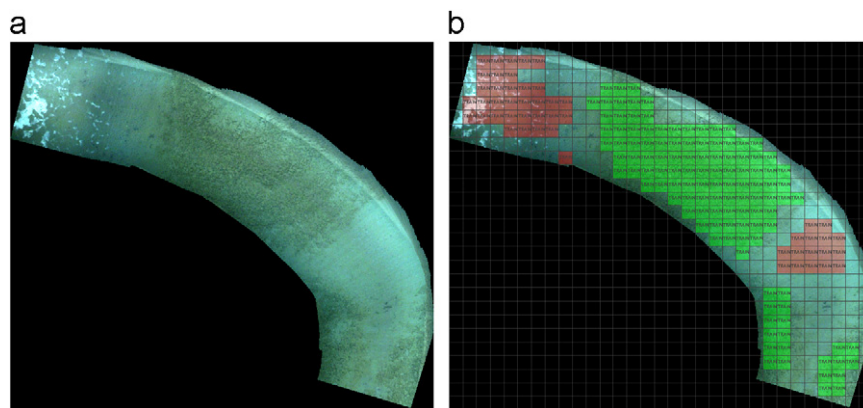
Fig. 3. (a) Sample seafloor video mosaic with Pogonophora presence in parts of the image data region. (b) Class labels manually assigned to grid cells during the training phase. Cells marked green indicate *presence* of Pogonophora coverage; cells marked red indicate *absence* of Pogonophora coverage. Cells not marked have not been annotated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
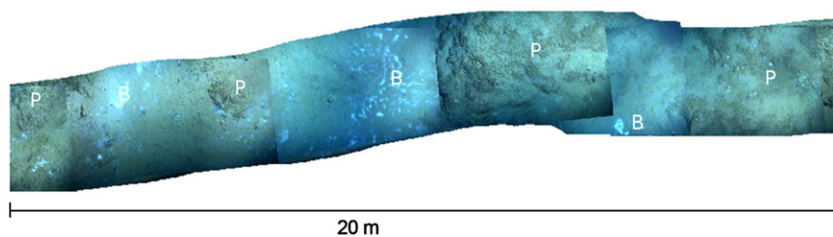
Fig. 4. Changing areas of tube worms and bacterial mats within a 20-m video mosaicking stripe northwest of the center of HMMV (*Beggiatoa* and Pogonophora samples marked with labels "B" and "P"). Five single georeferenced video mosaics have been concatenated by overlay applying a GIS.

and we expect the best classification performance here, other approaches are typically much faster in training. Thus, not only classification accuracy but also training speed are compared.

The following sections describe the image feature extraction, the preparation of training data, the selection of relevant features, and the process of selecting appropriate parameters for the machine learning schemes applied.

4.1. Image feature extraction

Automatic recognition of Pogonophora tube worms is achieved by learning visual properties of image regions with Pogonophora coverage described by numeric image features. Therefore, the data region of each image is partitioned into a regular grid (see Fig. 3). Image features are extracted and classification is performed based on cells. Classification is a two-class problem then: for each cell Pogonophora coverage is either *present* or *absent*.

A set of 49 numerical image features is extracted per cell describing visual low-level properties such as contrast, roughness and linelikeness. Image features are classical statistical and structural textural features (Tamura et al., 1978; Haralick et al., 1973; Haralick, 1979; Wu and Chen, 1992) widely applied for image classification tasks in different domains. Image features are listed with brief descriptions in Table 2. A cell size of 20×20 pixels

Table 1
Classifiers applied and parameter configurations tested per classifier.

Classifier	Parameters
Support vector machine (Vapnik, 1995)	Complexity constant C: $\log_2(C) = i, i = -5, -3, \ldots, 15$ RBF (radial basis function) kernel with γ: $\log_2(\gamma) = i, i = -15, -13, \ldots, 3$ (110 runs per set)
K-nearest-neighbors (Aha and Kibler, 1991)	Number of neighbors k: $k = 2, 5, 10, 20, 50, 100, 500, 1000$ With/without inverse distance weighting Search algorithms: Brute force (linear), Ball tree (Omohundro, 1989; Uhlmann, 1991), Cover tree (Beygelzimer et al., 2006), KD tree (Friedman et al., 1977) (64 runs per set)
C4.5 decision tree (Quinlan, 1993)	With/without pruning With/without Laplacian smoothing of predicted probabilities Minimum instances per leaf m: $m = 1, 2, \ldots, 20$ Pruning confidence threshold c: $c = 0.05, 0.1, \ldots, 0.5$ (440 runs per set)
Naive Bayes (John and Langley, 1995)	With kernel density estimator or normal distribution for numeric attributes With/without supervised discretization of numeric attributes for normal distribution (3 runs per set)

Table 2
Image features extracted per grid cell.

| Image feature	Description
f_1	Minimum gray level
f_2	Maximum gray level
f_3	Mean gray level
f_4	Number of gray levels used
f_5	Gray level standard deviation
f_6	*Polarization* of gray-level distribution, Tamura et al. (1978)
f_7	Texture *contrast*, Tamura et al. (1978)
f_8	Texture *directionality*, Tamura et al. (1978)
f_9	Texture *linelikeness*, Tamura et al. (1978)
f_{10}	*Roughness* of texture, as defined in Wu and Chen (1992)
$f_{11,12,13}$	*Contrast* computed from GLCM (mean|min|max over all directions, see below; same in the following), Haralick et al. (1973)
$f_{14,15,16}$	*Variance* computed from GLCM, Haralick et al. (1973)
$f_{17,18,19}$	*Entropy* computed from GLCM, Haralick et al. (1973)
$f_{20,21,22}$	*Correlation* computed from GLCM, Haralick et al. (1973)
$f_{23,24,25}$	*Angular second moment* computed from GLCM, Haralick et al. (1973)
$f_{26,27,28}$	*Inverse difference moment* computed from GLCM, Haralick et al. (1973)
$f_{29,30,31}$	*Sum average* computed from GLCM, Haralick et al. (1973)
$f_{32,33,34}$	*Sum variance* computed from GLCM, Haralick et al. (1973)
$f_{35,36,37}$	*Sum entropy* computed from GLCM, Haralick et al. (1973)
$f_{38,39,40}$	*Difference variance* computed from GLCM, Haralick et al. (1973)
$f_{41,42,43}$	*Difference entropy* computed from GLCM, Haralick et al. (1973)
$f_{44,45,46}$	*Information measure of correlation I* computed from GLCM, Haralick et al. (1973)
$f_{47,48,49}$	*Information measure of correlation II* computed from GLCM, Haralick et al. (1973)

Notes: In case of directed features extracted from GLCM (gray-level co-occurrence matrix, directions : $0°$, $45°$, $90°$, $135°$; Haralick et al., 1973) we use the mean, maximum, and minimum feature values over the four directions.

(or $\approx 18 \times 18$ cm for the given camera aperture and height over ground) was chosen for images of the underlying data set. This allows the computation of meaningful statistical image features while preserving good localization of detections. Many of the numerical image features are calculated on the basis of co-occurring gray levels expressed in terms of a so-called gray-level co-occurrence matrix (GLCM, Haralick et al., 1973; Haralick, 1979). For 8-bit input images this matrix has dimensions of 256×256. With decreasing cell size this matrix becomes sparsely populated up to a point where computation of statistical features becomes useless. On the other hand, with increasing cell-size classes tend to mix within cells (see example in Fig. 5c) leading to ambiguities. Further note that the models trained for recognition are specific to the imaging system as described in Section 3, especially regarding pixel scale.

4.2. Preparation of training data

The complete data set (see Section 3) consists of 4108 georeferenced video mosaics in total. By partitioning the image data regions into cells (see Fig. 3b), we get 632,350 cells, so the average number of cells per image is ≈ 154. For the training process 25,020 cells have been labeled manually with class labels *Pogonophora_present* and *Pogonophora_absent*. This complies with $\approx 4\%$ of the data set or ≈ 162 images if completely annotated.

The labeling of cells was performed by a human expert. There was no restriction and also no automatic support in the selection of sample images or cells during manual annotation, leaving these decisions purely up to the human expert and leading to a mostly random distribution of training samples across the investigation area. 2047 cells ($\approx 8.2\%$) have been labeled with *Pogonophora_present*, while 22,973 ($\approx 91.8\%$) have been labeled with *Pogonophora_absent*. The process took less than 1 day.

4.3. Feature selection

The presence of irrelevant or redundant features can have a negative impact on machine learning schemes (Witten and Frank,

2005). Therefore, a subset of features is often selected prior to learning. While support vector machines (Vapnik, 1995) are known to be robust regarding classification accuracy, the naive Bayes classifier of John and Langley (1995) is an example of a learning scheme that suffers from correlation of input features. Furthermore, besides classification accuracy, the training speed is also affected.

There are two classes of approaches commonly applied to feature subset selection. While *scheme-independent* approaches reduce the dimensionality of the feature space based on general characteristics of the data, *scheme-dependent* approaches (often referred to as "wrapper" methods) employ the specific machine learning scheme in that process. In this study, we apply the correlation-based, *scheme-independent* method introduced in Hall and Smith (1998) for feature subset selection, as

1. Hall and Smith (1998) have shown its equal performance compared to "wrapper" methods on standard machine learning data sets while being much faster. Moreover,
2. by application of a *scheme-independent* method for feature selection, it is ensured that the subset of selected features is the same for all classifiers in the subsequent experiments.

To test whether the classifiers suffer from irrelevance or redundancy of features, we perform two experiments. One is based on the full feature set and the other on a subset of selected features. The subset of six selected features is {f_1: minimum gray level, f_{20}: mean *correlation* (Haralick, 1979), f_{21}: maximum *correlation* (Haralick, 1979), f_{26}: mean *inverse difference moment* (Haralick, 1979), f_{27}: maximum *inverse difference moment* (Haralick, 1979), f_{10}: *roughness* of texture (Wu and Chen, 1992)}.
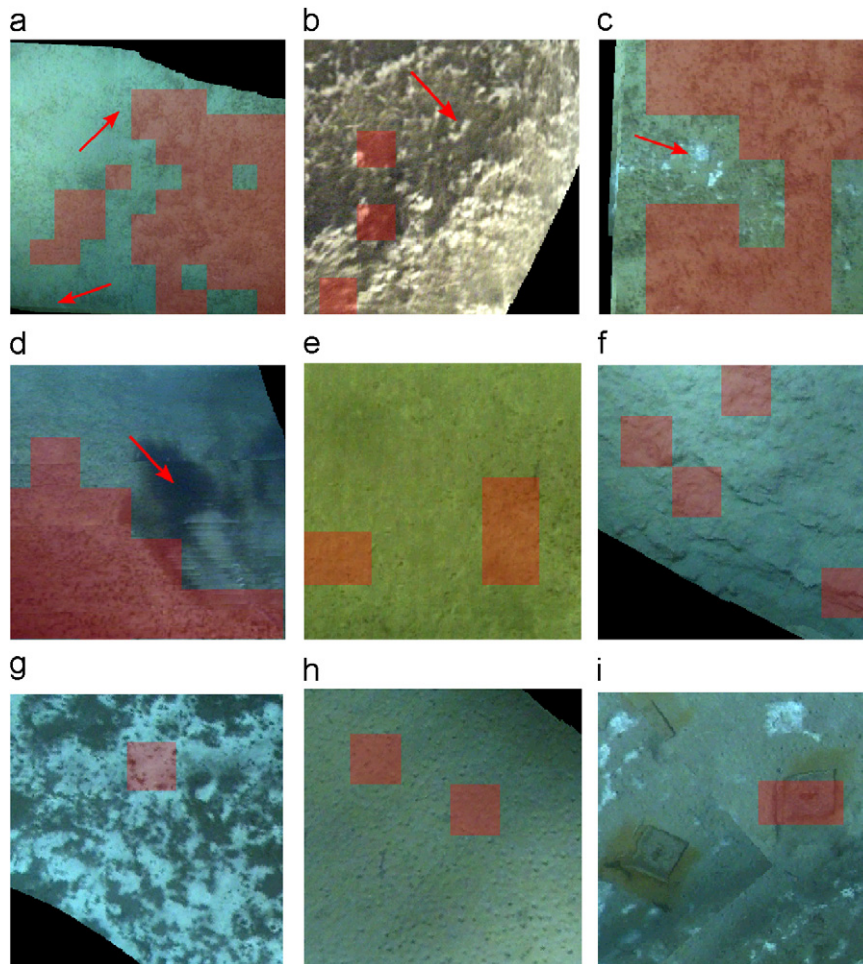
Fig. 5. Examples of missing or false detections (colored cells indicate detections, (a)–(d) missing detections, arrows point to undetected Pogonophora; (e)–(i) all colored cells false detections). (a) Undetected sparse Pogonophora coverage, (b) overexposure causing undetected Pogonophora coverage, (c) missing detections due to *Beggiatoa* mixing with Pogonophora inside cells, (d) missing detections due to mosaicking artifacts, (e) misdetections inside smooth mud, (f) misdetections at mud structures, (g) misdetections inside *Beggiatoa* field, (h) misdetections at small holes in sediment, (i) false detections due to unknown objects at the seafloor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.4. Parameter selection

Selecting optimal or near-optimal parameters is crucial for good performance of machine learning classifiers. The usual approach to parameter optimization is grid like exploration of the parameter space (Witten and Frank, 2005). Classifiers have been tested in various configurations listed in Table 1. For each configuration, a stratified 10-fold cross validation was performed on the training data and the models were ranked by overall (average) accuracy. In cases of multiple best configurations with equal accuracy, the fastest in training was chosen.

5. Results

In this section, results obtained by application of the different machine learning classifiers (see Table 1) to the training data are compared. Table 3 lists the best classification rate achieved per classifier using a set of parameters selected as described in Section 4.4 for both feature sets. The best classification rate of 98.86% has been achieved by application of support vector machines (Vapnik, 1995) to the full feature set. These also perform best on the set of selected features.

The K-nearest-neighbors (Aha and Kibler, 1991) and C.45 decision tree classifier (Quinlan, 1993) perform nearly equal. As support vector machines, they perform slightly better on the full feature set. The naive Bayes classifier (John and Langley, 1995) obviously suffers from correlation of features in the full set. This results in much better classification performance on the set of selected features, while still performing worse than the other classifiers.

Support vector machines performed less well than the other classifiers regarding training speed, whereas the training time is still acceptable for the targeted application. The best choices in speed are the K-nearest-neighbors and C.45 decision tree classifiers.

Using the best model on the training data (support vector machines applied to the full feature set, see Table 3) the complete image data set has been analyzed. The results are visualized in Fig. 6.

6. Discussion

In the following sections we further discuss the results of the proposed approach. Section 6.1 presents tests based on a different reference data set, where larger regions have been annotated manually by a domain expert. While being annotated more roughly in terms of exact localization of seafloor features, this data set covers much more of the overall investigation area than

Table 3
Best classification rates achieved per classifier applied to the full feature set and the set of selected features.

Classifier	Correctly classified	False positives	False negatives	Time (s)
All features				
Support vector machine[a]	24,734	118	168	10,871
	(98.86%)	(0.47%)	(0.67%)	
K-nearest-neighbors[b]	24,498	131	391	83
	(97.91%)	(0.52%)	(1.56%)	
C.45 decision tree[c]	24,469	186	365	40
	(97.8%)	(0.74%)	(1.46%)	
Naive Bayes[d]	19,833	307	4880	473
	(79.27%)	(1.23%)	(19.5%)	
Selected features				
Support vector machine[e]	24,201	229	590	14,583
	(96.73%)	(0.92%)	(2.36%)	
K-nearest-neighbors[f]	24,148	261	611	13
	(96.51%)	(1.04%)	(2.44%)	
C.45 decision tree[g]	24,119	323	578	4
	(96.4%)	(1.29%)	(2.31%)	
Naive Bayes[h]	23,807	555	658	1
	(95.15%)	(2.22%)	(2.63%)	

[a] Vapnik (1995), $\log_2(C) = 11$, RBF kernel with $\log_2(\gamma) = -3$.
[b] Aha and Kibler (1991), $k = 10$, inverse distance weighting, *Cover Tree* search (Beygelzimer et al., 2006).
[c] Quinlan (1993), $m = 16$, with pruning, $c = 0.1$, with Laplacian smoothing.
[d] John and Langley (1995), with kernel density estimator.
[e] Vapnik (1995), $\log_2(C) = 7$, $\log_2(\gamma) = 3$.
[f] Aha and Kibler (1991), $k = 20$, inverse distance weighting, *Ball Tree* search (Omohundro, 1989; Uhlmann, 1991).
[g] Quinlan (1993), $m = 10$, with pruning, $c = 0.3$, without Laplacian smoothing.
[h] John and Langley (1995), normal distribution, without supervised discretization.
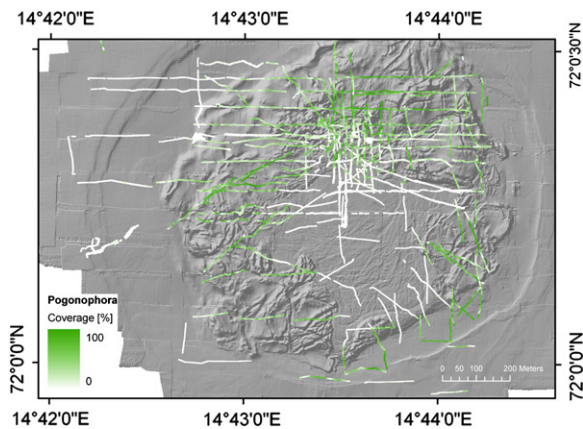
Fig. 6. Spatial distribution of Pogonophora coverage obtained by automatic image analysis of the complete data set (all 4108 video mosaics). Overlain on the HMMV bathymetry visualized as a hill-shading map.

the training data. This is followed by a detailed analysis of the causes of misdetections in Section 6.2.

6.1. Quality assessment by comparison with visually inspected field data

The data we refer to in further tests described in this section is a shape file consisting of 1578 polygons covering $\approx 60\%$ of the overall image data (in contrast to only $\approx 4\%$ for the training data). A single polygon does not necessarily correspond to the data region of an image, but typically covers a larger area. Polygons cover regions of similar density regarding Pogonophora coverage.

The following coverage classes have been assigned manually to polygons of the reference data set: *No Pogonophora coverage*, *less than*20% *Pogonophora coverage*, 20–50% *Pogonophora coverage*, 50–80% *Pogonophora coverage*, and *more than*80% *Pogonophora coverage*. The reference data set covers an area of $\approx 45,994$ m^2 and is visualized in Fig. 7a.

This data set covers much more of the investigation area and implicitly contains information on the local neighborhoods of detections. It allows a better assessment of

- the degree of generalization of the models learned (is a model performing $\approx 1-2\%$ better in a cross validation on the training data really better or does the difference in measured classification rates result from pure overfitting to the training data?), and
- the influence of density of Pogonophora coverage (are there more misdetections in, e.g., very densely or sparsely covered areas?).

Furthermore, considering larger polygons with approximately equal Pogonophora coverage instead of single cells, it was feasible to perform a statistical evaluation of the causes of misdetections.

To be able to compare the estimated coverage degree annotated by the domain expert with the cell-based detections, we need to cumulate detections across polygonal regions. The coverage degree for a single polygon is obtained as follows:

1. From the set of all intersection segments of image data regions with the reference polygon we compute a set of *atomic* segments associated with a distinct set of image regions. (Assume a reference polygon covering two intersecting image data regions *A* and *B*. This would result in three *atomic* segments, one containing the part of the intersection of image data region *A* with the reference polygon excluding the intersection area with the data region *B*, the same for image data region *B*, and the intersection area of *A* and *B*.)
2. For each *atomic* segment we independently compute the coverage degree as the average ratio of all contained cells with detections to the overall number of contained cells from all associated image regions.
3. The coverage of the complete polygon is then computed as a combination of the coverage of *atomic* segments weighted by segment areas.

The comparison with reference polygons is then based on the *measurable* deviation from the annotated coverage interval. E.g., in the case of a polygon annotated with 50–80% *Pogonophora coverage* and a coverage degree of 45% obtained by automatic analysis, underestimation of 5% is assumed.

Deviations of results from automatic image analysis using the model that performed best on the training data (support vector machine applied to the full feature; see Table 1) compared to the reference polygons are visualized in Fig. 7b. For 1207 of the 1578 polygons there was no measurable error (green polygons in the visualization), while for the rest of the polygons there were mostly only minor deviations (near-white polygons).

Table 4 presents a more detailed analysis of measurable deviations for the best models of *all* classifiers tested and *both* underlying feature sets. The average measurable error is presented per coverage class and for the complete reference data set. Also, here, the average of errors among polygons are computed as a combination of single errors weighted by shape area. Again, the best result, with an average measurable error of 1.52%, was obtained by application of support vector machines to the full feature set. Error rates here are also the most constant across different coverage classes, while the other classifiers tend to

produce higher errors in more densely covered areas. For all classifiers except naive Bayes (see Section 4.4), results based on the full feature set are better on average. Performance of the K-nearest-neighbors classifier and C.45 decision trees is again comparable. In contrast to the results on the training data, the C.45 classifier performs slightly better. As on the training data, the Naive Bayes classifier performs worse than the other classifiers tested on both feature sets. In general, the tendencies obtained in the tests on the training data are approved. This leads to the conclusion that the better classification rates result from better generalization rather than from overfitting to the training data.

6.2. Causes of misdetection

While error rates obtained by automatic image analysis are generally low (see Sections 5 and 6.1), there are still misdetection. To further identify the causes of misdetections, another experiment was carried out. For all polygons where a deviation to the reference data set occurred (see Section 6.1 and Fig. 7b), images and detections have been revised and misdetections have been assigned to error classes. Again, results from the best-performing model (support vector machines applied to the full feature set) were used as reference results for automatic image analysis. Obviously, in these tests we might miss a small number of errors in cases where the domain expert and the automatic image analysis produced the same false estimation, but since a reasonably small error rate can be assumed for both, such cases are very unlikely.

The resulting error classes and the fractions of the overall error caused by misdetections of the respective classes are listed in Table 5. Fig. 5 displays example images.

7. Conclusions

Related to the investigation of methane discharge at the Håkon Mosby Mud Volcano, an approach to automatic detection of Pogonophora coverage in georeferenced seafloor video mosaics was developed. The recognition is based on textural features extracted on the basis of cells obtained by partitioning the mosaic data regions into a regular grid. For classification, four state-of-the art machine learning schemes have been applied, with classification rates up to 98.86% in a cross validation on the training data. Further tests based on a larger reference data set covering $\approx 60\%$ of the overall investigation area revealed an average measurable error of $\approx 1.52\%$.

Means of automatic image analysis are essential to cope with the increasing volume of georeferenced image data acquired by mobile underwater platforms such as ROVs and AUVs. These enable scientists to gain insight into newly acquired data without time-consuming manual analysis. Moreover, such means of automatic ad hoc analysis can be particularly useful for operations planning during an ongoing expedition or on fully autonomous platforms.
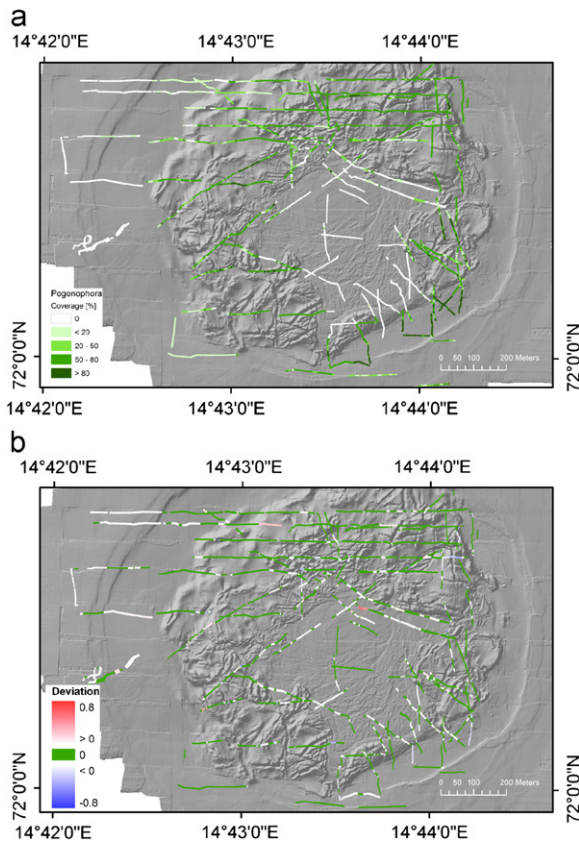
Fig. 7. (a) Coverage degree obtained through visual inspection. (b) Deviations of result from automatic image analysis. Green: no measurable difference (automatically detected Pogonophora coverage lies in the range of the coverage class annotated by the domain expert). White to red: Pogonophora coverage was overestimated. White to blue: Pogonophora coverage was underestimated. (Both overlain on the HMMV bathymetry visualized as a hill-shading map.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Average measurable error per coverage class (see Section 6.1) and overall.

Classifier	Coverage class					Overall
	None	$< 20\%$	**20–50%**	**50–80%**	$> 80\%$	
All features (avg. measurable error [%])						
Support vector machine	1.40	0.77	0.83	2.45	1.44	1.52
K-nearest-neighbors	1.06	0.54	1.52	4.93	5.62	2.64
C.45 decision tree	1.16	0.51	1.73	4.62	4.25	2.46
Naive Bayes	13.33	7.86	5.69	3.36	2.53	7.32
Selected features (avg. measurable error [%])						
Support vector machine	0.62	0.14	4.68	10.96	13.97	5.52
K-nearest-neighbors	0.55	0.13	4.44	11.76	15.56	5.86
C.45 decision tree	0.93	0.23	3.75	8.75	10.52	4.49
Naive Bayes	1.58	0.41	3.17	8.92	12.95	4.94
Area (m^2)						
–		13,719	7632	6512	13,164	4967	45,994

Table 5
Classes of misdetections and fractions of the overall error.

Error class	Error (%)
Causes of underestimation	
[U1] Sparse Pogonophora coverage has partly not been detected. From the visual impression this is typically very close to smooth mud. Fig. 5a displays an example. In comparison to Fig. 5e you can see how small the difference from smooth mud is in certain regions.	34.12
[U2] In rare cases images with overexposure occur, leading to significant undetected Pogonophora coverage (see Fig. 5b). This is probably due to the ROV being too close to the seafloor (also, the scale in the world files indicates that)	6.90
[U3] Regions occur where Pogonophora mix with *Beggiatoa* inside cells (see Fig. 5c). In such cases feature values can be influenced to a degree leading to undetected Pogonophora	5.96
[U4] Pogonophora coverage has party not been detected due to mosaicking artifacts (see Fig. 5d)	1.95
Causes of overestimation	
[O1] No false detections could be identified. Such errors are caused by the fact that cells classified as Pogonophora_present are not 100% covered (see Figs. 5a and b). While the human annotator tends to take this into account, this cannot be distinguished by the automatic detection, leading to an overestimation	23.05
[O2] Pogonophora coverage has been detected inside smooth mud. This can be regarded as the counterpart of class [U1]. Fig. 5e displays an example	10.80
[O3] False detections at mud structures or relief. Fig. 5f displays an example	8.41
[O4] Rarely occurring false detections at *Beggiatoa*. Fig. 5g displays an example	4.66
[O5] Misdetection at small holes in sediment. See Fig. 5h for an example	3.15
[O6] Misdetections due to mosaicking artifacts. Such artifacts can cause both false and missing detections. Fig. 5d displays an example of missing detections	0.27
[O7] Misdetection caused by objects seldom occurring (e.g., fishes, waste) and therefore not part of the training set. Fig. 5i displays an example of waste on the seafloor	0.21
Special cases	
Annotated region or *atomic* segments too small. None or few cells fit	0.52

References

Aha, D., Kibler, D., 1991. Instance-based learning algorithms. Machine Learning 6, 37–66.

Allais, A.-G., Borgetto, M., Opderbecke, J., Pessel, N., Rigaud, V., 2004. Seabed video mosaicing with MATISSE: a technical overview and cruise results. In: Proceedings of the 14th International Offshore and Polar Engineering Conference, vol. 2, Toulon, France, pp. 417–421.

Beygelzimer, A., Kakade, S., Langford, J., 2006. Cover trees for nearest neighbor. In: ICML'06: Proceedings of the 23rd International Conference on Machine Learning. ACM Press, New York, NY, USA, pp. 97–104.

Boetius, A., Ravenschlag, K., Schubert, C.J., Rickert, D., Widdel, F., Gieseke, A., Amann, R., Joergensen, B.B., Witte, U., Pfannkuche, O., 2000. A marine microbial consortium apparently mediating anaerobic oxidation of methane. Nature 407, 623–626.

Damm, E., Budéus, G., 2003. Fate of vent-derived methane in seawater above the Håkon Mosby Mud Volcano (Norwegian Sea). Marine Chemistry 82, 1–11.

de Beer, D., Sauter, E., Niemann, H., Witte, U., Boetius, A., 2006. In situ fluxes and zonation of microbial activity in surface sediments of the Håkon Mosby Mud Volcano. Limnology and Oceanography 51 (3), 1315–1331.

Feseker, T., Foucher, J.-P., Harmegnies, F., 2008. Fluid flow or mud eruptions? Sediment temperature distributions on Håkon Mosby mud volcano, SW Barents Sea slope. Marine Geology 247 (3–4), 1–14.

Friedman, J.H., Bentley, J.L., Finkel, R.A., 1977. An algorithm for finding best matches in logarithmic expected time. ACM Transactions on Mathematical Software 3 (3), 209–226.

Gebruk, A., Krylova, E., Lein, A., Vinogradov, G., Anderson, E., Pimenov, N., Cherkashev, G., Crane, K., 2003. Methane seep community of the Håkon Mosby mud volcano (the Norwegian Sea): composition and trophic aspects. Sarsia: North Atlantic Marine Science 88 (6).

Guinan, J., Grehan, A.J., Dolan, M.F.J., Brown, C., 2009. Quantifying relationships between video observations of cold-water coral cover and seafloor features in Rockall Trough, west of Ireland. Marine Ecology Progress Series 375, 125–138.

Hall, M.A., Smith, L.A., 1998. Practical feature subset selection for machine learning. In: McDonald, C. (Ed.), Proceedings of the 21st Australasian Computer Science Conference ACSC'98, Springer, pp. 181–191.

Haralick, R.M., 1979. Statistical and structural approaches to texture. Proceedings of the IEEE 67 (5), 786–804.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics 3 (6), 610–621.

Hinrichs, K.U., Boetius, A., 2002. The anaerobic oxidation of methane: new insights in mircobial ecology and biogeochemistry. In: Wefer, G., Billett, D., Hebbeln, D., Joergensen, B., Schlüter, M., Weering, T.V. (Eds.), Ocean Margin Systems. Springer-verlag, Berlin, Heidelberg, pp. 457–477.

Hjelstuen, B.O., Eldholm, O., Faleide, J.I., 2007. Recurrent Pleistoceine mega-failures on the SW Barents Sea margin. Earth Planetary Science Letters 258, 605–618.

Hjelstuen, B.O., Eldholm, O., Faleide, J.I., Vogt, P.R., 1999. Regional setting of Håkon Mosby Mud Volcano, SW Barents Sea margin. Geo-marine Letters 19 (1–2), 22–28.

Jerosch, K., Lüdtke, A., Schlüter, M., Ioannidis, G.T., 2007. Automatic content-based analysis of georeferenced image data: detection of Beggiatoa mats in seafloor video mosaics from the Håkon Mosby Mud Volcano. Computers & Geosciences 33, 202–218.

Jerosch, K., Schlüter, M., Foucher, J.-P., Allais, A.G., Klages, M., Edy, C., 2006a. Spatial distribution of mud flows, chemoautotrophic communities, and biogeochemical habitats at Håkon Mosby Mud Volcano. Marine Geology 243, 1–17.

Jerosch, K., Schlüter, M., Pesch, R., 2006b. Spatial analysis of marine categorical information using indicator kriging applied to georeferenced video mosaics of the deep-sea Håkon Mosby Mud Volcano. Ecological Informatics 1, 391–406.

John, G.H., Langley, P., 1995. Estimating continuous distributions in Bayesian classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann, San Mateo, CA, pp. 338–345.

Kaul, N., Foucher, J.-P., Heesemann, M., 2006. Estimating mud expulsion rates from temperature measurements on Håkon Mosby Mud Volcano, SW Barents Sea. Marine Geology 229, 1–14.

Kohler, K.E., Gill, S.M., 2006. Coral point count with excel extensions (CPCe): a visual basic program for the determination of coral and substrate coverage using random point count methodology. Marine Geology 32, 1259–1269.

Kopf, A.K., 2002. Significance of mud volcanism. Reviews of Geophysics 40 (2).

Kostylev, V.E., Todd, B.J., Fader, G.B.J., Courtney, R.C., Cameron, G.D.M., Pickrill, R.A., 2001. Benthic habitat mapping on the Scotian Shelf based on multibeam bathymetry, surficial geology and sea floor photographs. Marine Ecology Progress Series 219, 121–137.

Lein, A., Vogt, P., Crane, K., Egorov, A., Ivanov, M., 1999. Chemical and isotopic evidence for the nature of the fluid in CH4-containing sediments of the Håkon Mosby Mud Volcano. Geo-marine Letters 19, 76–83.

Lösekann, T., Knittel, K., Nadalig, T., Fuchs, B., Niemann, H., Boetius, A., Amann, R., 2007. Diversity and abundance of aerobic and anaerobic methane oxidizers at the Håkon Mosby Mud Volcano, Barents Sea. Applied and Environmental Microbiology 73 (10), 3348–3362.

Lösekann, T., Robador, A., Niemann, H., Knittel, K., Boetius, A., Dubilier, N., 2008. Endosymbiosis between bacteria and deep-sea siboglinid tubeworms from an Arctic Cold Seep (Håkon Mosby Mud Volcano, Barents Sea). Environmental Microbiology 10 (12), 3237–3254.

Miene, A., Möhrke, O., 1997. Analyse und Beschreibung von Texturen. Master's Thesis, Universität Bremen.

Milkov, A.V., 2000. Worldwide distribution of submarine mud volcanoes and associated gas hydrates. Marine Geology 167, 29–42.

Milkov, A.V., Vogt, P.R., Cherkashev, G., 1999. Sea-floor terrains of Håkon Mosby Mud Volcano as surveyed by deep-tow video and still photography. Geo-marine Letters 19, 38–47.

Milkov, A.V., Vogt, P.R., Crane, K., Lein, A.Y., Sassen, R., Cherkashev, G., 2004. Geological, geochemical, and microbial processes at the hydrate-bearing Håkon Mosby Mud Volcano: a review. Chemical Geology 205, 347–366.

Niemann, H., Lösekann, T., de Beer, D., Elvert, M., Nadalig, T., Knittel, K., Amann, R., Sauter, E.J., Schlüter, M., Klages, M., Foucher, J.-P., Boetius, A., 2006. Novel microbial communities of the Håkon Mosby mud volcano and their role as a methane sink. Nature 443, 854–858.

Omohundro, S.M., 1989. Five Balltree Construction Algorithms. Technical Report TR-89-063, International Computer Science Institute.

Omoregie, E.O., Niemann, H., Mastalerz, V., de Lange, G.J., Stadnitskaia, A., Mascle, J., Foucher, J.-P., Boetius, A., 2008. Microbial methane oxidation and sulfate reduction at cold seeps of the deep Eastern Mediterranean Sea. Marine Geology 261, 114–127.

Perez-Garcia, C., Feseker, T., Mienert, J., Berndt, C., 2009. The Håkon Mosby Mud Volcano: 330 000 years of focused fluid flow activity at the SW Barents Sea slope. Marine Geology 262, 105–115.

Pimenov, N., Savvichev, A., Rusanov, I., Lein, A., Egorov, A., Gebruk, A., Moskalev, L., Vogt, P., 1999. Microbial processes of carbon cycle as the base of food chain of Håkon Mosby Mud Volcano benthic community. Geo-marine Letters 19, 86–196.

Purser, A., Bergmann, M., Lundälv, T., Ontrup, J., Nattkemper, T., 2009. Use of machine-learning algorithms for the automated detection of cold-water coral habitats—a pilot study. Marine Ecology Progress Series 397, 241–251.

Quinlan, R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo, CA.

Sauter, E.J., Muyakshin, S.I., Charlou, J.-L., Schlüter, M., Boetius, A., Jerosch, K., Damm, E., Foucher, J.-P., Klages, M., 2007. Methane discharge from a deep-sea submarine mud volcano into the upper water column by gas hydrate-coated methane bubbles. Earth and Planetary Science Letters 243 (3–4), 354–365.

Smirnov, R.V., 2000. Two new species of Pogonophoran from the arctic mud volcano off northwestern Norway. Sarsia 85, 141–150.

Soltwedel, T., Portnova, D., Kolar, I., Mokievsky, V., Schewe, I., 2005. The small-sized benthic biota of the Håkon Mosby Mud Volcano (SW Barents Sea slope). Journal of Marine Systems 55, 271–290.

Tamura, H., Mori, S., Yamawaki, T., 1978. Texture features corresponding to visual perception. IEEE Transactions on Systems, Man, and Cybernetics 8 (6), 460–473.

Uhlmann, J.K., 1991. Satisfying general proximity/similarity queries with metric trees. Information Processing Letters 40 (4), 175–179.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer-verlag.

Vincent, A.G., Jouffroy, J., Pessel, N., Opderbecke, J., Borgetto, M., Rigaud, V., 2003. Real-time georeferenced video mosaicing with the MATISSE system. In: Proceedings of the Oceans 2003 Marine Technology and Ocean Science Conference, MTS/IEEE OCEANS'03, vol. 4, San-Diego, USA, pp. 2319–2324.

Vogt, P., Cherkashev, G., Ginsburg, G., Ivanov, G., Milkov, A., Crane, K., Lein, A., Sundvor, E., Pimenov, N., Egorov, A., 1997. Håkon Mosby Mud Volcano provides unusual example of venting. EOS 78 (549), 556–557.

Vogt, P.R., Gardner, J., Crane, K., 1999. The Norwegian-Barents-Svalbard (NBS) continental margin: introducing a natural laboratory of mass wasting, hydrates and ascent of sediment pore water and methane. Geo-Marine Letters 19, 2–21.

Witten, I.H., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques, second ed. Morgan Kaufmann, San Francisco, CA, USA.

Wu, C.-M., Chen, Y.-C., 1992. Statistical feature matrix for texture analysis. CVGIP: Graphical Models and Image Processing 54 (5), 407–419.

Zhanga, Q., Pavlica, G., Chena, W., Frasera, R., Leblanca, S., Cihlara, J., 2005. A semi-automatic segmentation procedure for feature extraction in remotely sensed imagery. Computers & Geosciences 32, 289–296.