Investigation of hidden parameters influencing the automated object detection in images from the deep seafloor of the HAUSGARTEN observatory

Timm Schoening (tschoeni@cebitec.uni-bielefeld.de), Biodata Mining Group, Faculty of Technology, Bielefeld University
Melanie Bergmann, HGF-MPG Group for Deep-Sea Ecology and Technology,
Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany
Antje Boetius, Max Planck Institute for Marine Microbiology, Bremen, Germany
Tim W. Nattkemper, Biodata Mining Group, Faculty of Technology, Bielefeld University

Abstract:
Detecting objects in underwater image sequences and video frames automatically, requires the application of selected algorithms in consecutive steps. Most of these algorithms are controlled by a set of parameters, which need to be calibrated for an optimal detection result. Those parameters determine the effectivity and efficiency of an algorithm and their impact is usually well known. There are however further non-algorithmic impact factors (or hidden parameters), which bias the training of a machine learning system as well as the subsequent detection process and thus need to be well understood and taken into account.
In benthic imaging, one dominant, hidden parameter is the distance of the image acquisition device above the seafloor. Variations in the distance lead to variations in the benthic area size being captured, the relative size and position of an object within an image, the effect of the artificial light source and thus the recorded color spectrum. Image processing techniques that allow modeling the induced variations can be used to compensate for those effects and thus allow the exploration of initially biased data. Those processing techniques again require algorithmic parameters, which are influenced by the hidden parameters contained within the initial data.
In supervised machine-learning architectures, further challenges arise from the inclusion of human expert knowledge used for the training of the learning algorithm. Utilizing the knowledge of only one expert can conceal the information needed for the generalization capability of an automated semantic image annotation system. Utilizing the knowledge of several experts requires explicit instruction of the participants to be able to produce comparable results. The fusion of individual expert knowledge poses further hidden parameters that impact the supervised learning architecture. Those could be an individual object specific expertise or the tendency to annotate with more or less self-criticism, which together can be expressed as the expert's trustworthiness.
In the context of mega-fauna detection in benthic images, we investigate the effects of some of these parameters on our machine learning based detection system iSIS that consists of four succeeding steps: Imaging, expert annotation, training, and detection. The images to be analyzed were taken at the deep-sea, long-term observatory HAUSGARTEN and five experts created an annotation gold standard.
We found, that the hidden parameters from imaging as well as the fusion of expert knowledge could partly be compensated and were able to achieve detection performances of 0.67 precision and 0.87 recall. Despite the efforts to compensate the hidden parameters, the detection performance was still varying across the image transect. This poses the potential occurrence of further hidden parameters not taken into account so far.

Introduction:

Marine research institutions and commercial companies around the world are capturing benthic images to monitor various processes from gas spills to impacts on habitats caused by environmental change. The evaluation of the vast amount of image data constitutes a bottleneck as manual evaluation is time-consuming, error-prone [1] and automated approaches are still under development and usually suited to fit the needs of one distinct research project [2,3,4].

One of these projects is the long-term observatory Hausgarten [5] in the North Pacific where benthic mega fauna taxa are monitored over time. So far, the individual taxa are hand labeled by experts in the browser-based annotation software BIIGLE [6]. To extend the amount of mega fauna detections without the need of further experts or more of their time, a supervised machine learning approach, called iSIS (intelligent Screening of Image Series) was developed [7]. This approach is based on the currently available taxa (i.e object) annotations and was applied to unseen data for classification and automated generation of further annotations [8].

Utilizing the additional semantic information, the data vault of multi-year, multi-location image data will be opened for analysis by biologists. Image processing and machine learning can then be used as an additional tool for habitat mapping, species interaction analysis and change monitoring.

During the automation of the detection process several computational steps were required to create comparable results. The images had to be pre-processed to equalize the lighting conditions and color spectra. This preprocessing was tuned utilizing the expert annotations. Multimodal feature extraction was applied at object positions and individual Support Vector Machines [9] were trained for each object type to learn its feature representation. Those SVMs were then used to classify feature extractions of the whole image. From these classification results, automated detection positions of objects were derived that could be compared to the hand labeled object positions. iSIS was tuned to create an optimal result for all object type combined. The complete setup of iSIS is shown in figure 1.

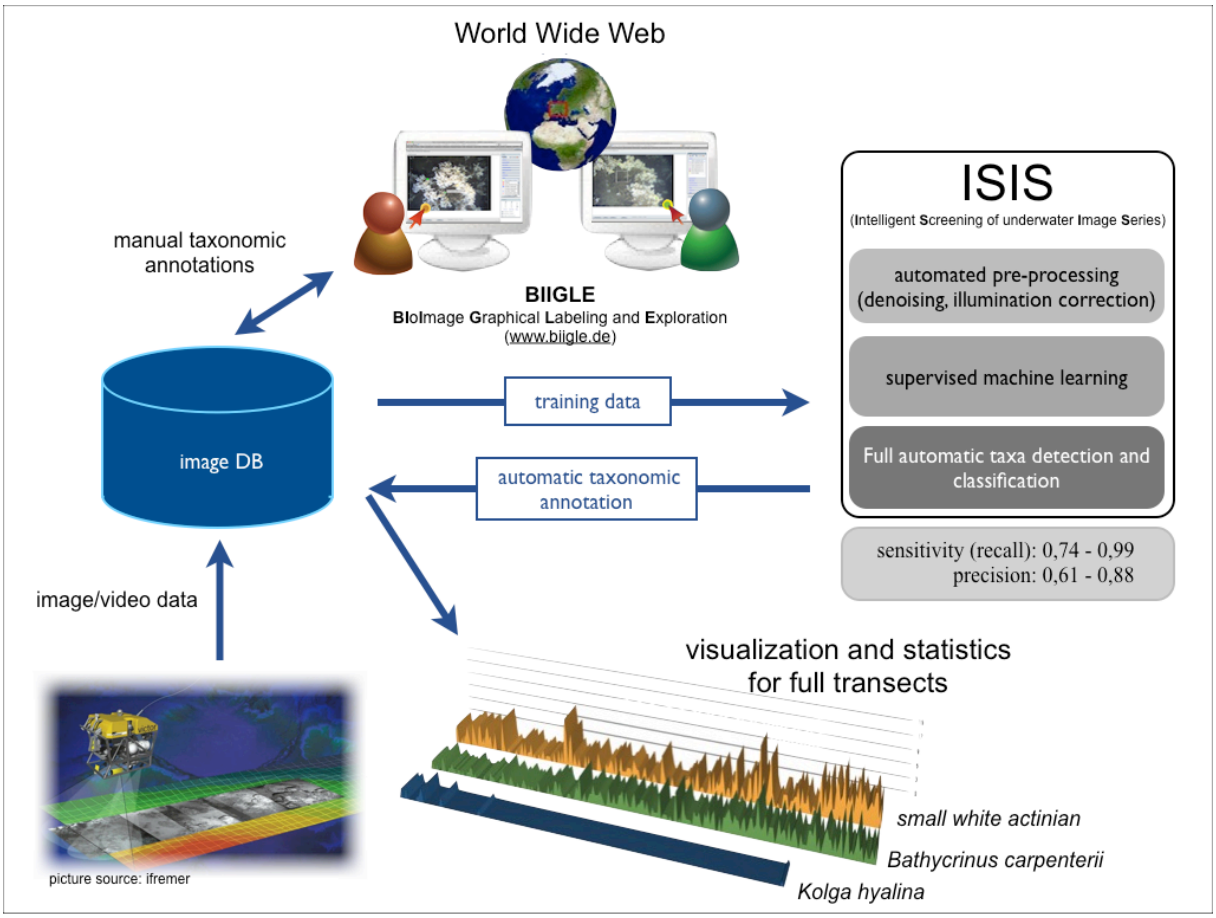

Figure 1: Schematic overview of the automated object detection process. Via the online image annotation software BIIGLE, expert knowledge is gained (top) and stored in a database, together with the images (middle left) that were recorded by an OFOS (bottom left). From those annotations, training data is created and used for the machine-learning step of iSIS (right). The trained SVMs are then used to find further occurrences within previously unseen images from the database. Those occurrence counts can then be visualized over the length of a transect (bottom center).

The different steps that make up iSIS, require a set of parameters (e.g. Gaussian kernel size, SVM confidence thresholds) that are well known and implicitly required to execute the algorithms. However, during the evaluation of specific components, some hidden parameters were considered or discovered, that had a varying impact on subsequent steps of the automated detection.

Here, three impact factors will be described: the area of the image in square meters, the position of an object within the image and the consensus of human experts on the occurrence of an object. While the image area was anticipated as an influencing parameter and dealt with by means of the illumination correction, the other two parameters arose during the optimization. They showed to pose an opportunity for further tuning of the automated detection system.

Materials:

An image transect, taken 2004 at station IV of Hausgarten was used to tune the detection system. A remotely controlled Ocean Floor Observation System (OFOS) took the images and projected three laser markers on the seafloor that allow to compute the area in square meters, covered by an image.

Five experts hand labeled the positions of seven distinct object types within 70 images by a pixel coordinate. Thus, each hand label can be described by *l(t,u,i,x,y)* where *t* defines the object type (*t* elem {1..7}), *u* specifies the expert (*u* elem {1..5}), *i* is the index of the image (*i* elem {1..70}) and *x,y* define the position of the label within the image. On average, 40 different object instances were labeled per image.

Labels of the same type within close vicinity were fused to gold labels *g(t,s,i,x,y)* with *t* and *i* as before. *x,y* are computed as the centroid of the x and y positions of all fused labels and *s* gives the supporter count which is the amount of labels *l* that were fused to one gold label *g* (*s* elem {1..5}).

Methods:
The original images cover a variety of color spectra and illumination conditions. Initial efforts for the automated detection resulted in poor training and test performances. The images thus had to be preprocessed to be comparable. A Gaussian Filter with a large kernel (701x701 pixel) was applied and the result subtracted from the original image. The histograms of the so equalized images were then non-linearly transformed to place the peak of the intensity histogram to the middle of the color range (i.e.127).

The lightness peak within the Gaussian filtered image was used as a reference regarding the location of a label within the image. The peak can be described by *p(i,x,y)* where *i* is the index of the image, and *x* and *y* define the position of the peak within the image. The distance of a label *l(t,u,i,x,y)* to the peak *p(i,x,y)* was computed as the Euclidean distance: *d = sqrt(pow(l_x-p_x,2) + pow(l_y-p_y,2))*. Those continuous distance values were assigned to one of eleven equally distributed distance bins *b* (*b* elem {1..11}). A gold label can thus be described by *g(t,s,l,b)*, omitting the more detailed *x,y*-position.

Gold labels with a supporter count of at least 3 were used as positive training samples for the objects. It was assumed that objects, which gathered a supporter count below three, are untrustworthy and were thus neglected. At all training positions, feature vectors were extracted that consisted of MPEG7 descriptors as well as Gabor responses to cover color, structure and texture features together. This resulted in a multimodal feature representation with 424 dimensions for each pixel.

SVMs were trained for each object type individually. The training sets for each of the seven SVM consisted of 50% samples of the SVM's object and 25% feature vectors representing all other object types in equal proportions. The remaining 25% were samples of the sediment background. The tuning of the SVM parameters was done with four-fold cross-validation while one tenth of the images were removed from training as well as the cross validation to allow for generalization estimation.

The trained SVMs were then used to classify the feature representations of every pixel in all seventy images, yielding seven object specific confidence values for each pixel. A tuned threshold binarized those confidence values. The centroids of connected regions within the confidence maps were used as the automated detection positions and were compared to the expert labels. Thus, a further attribute, the performance *p*, can be assigned to each label, where those that have been correctly identified are counted as True Positives (*p*=TP), annotations that were overlooked are counted as False Negatives (*p*=FN) and detections for

which no annotation was found are counted as False Positives (*p*=FP). From these numbers, precision *(Sum TP/(Sum TP + Sum FP))*, recall *(Sum TP/(Sum TP + Sum FN))* and the F1-rate *(2*TP/(2*TP+FP+FN)* were computed as measures for the detection performance regarding a group of labels.

To see, whether we dealt with all hidden parameters, influencing the detection process, we conducted studies on different groups of labels, characterized by the label attributes (type *t*, image *i*, supporter *s*, bin *b* and performance *p*).

At first we correlated the F1-rate of all labels (independent of *t*, *s* and *b*) to the area of seafloor covered by the image they were marked in. A successful preprocessing should prevent a correlation between the F1-rate and the image area. Also, we performed the same experiment with a distinction between label types.

As we found implications of further hidden parameters during the evaluation of individual object detections, we correlated and visualized other combinations of label attributes (e.g. recall vs. *s*, *b* vs. label density).

To see, whether the experts marked items based on their position within the image, we grouped all labels independent of *t*, *s* and *i* and investigated the label density per *b*. Therefore we computed the area of each bin of each image, as this value is dependent on the position of the peak within the image.

Results:
During the expert annotation, a total of 13,699 annotations were gathered. Those were fused to 2,634 gold labels. The total detection performance was 0.67 (precision) and 0.87 (recall). The object-specific numbers are given in table 1.

Object type	**# labels**	**# gold labels**	**Recall**	**Precision**	**F1-rate**
Bathycrinus carpenterii	2524	503	0.74	0.61	0.67
Bathycrinus stalks	1729	341	0.63	0.38	0.47
Burrow	5701	1112	0.93	0.50	0.65
Purple anemone	498	97	0.69	0.28	0.40
Kolga hyalina	172	30	1.00	0.88	0.94
White anemone	2438	457	0.86	0.60	0.71
White sponge	637	94	0.89	0.43	0.58
Total	**13699**	**2634**	**0.87**	**0.67**	**0.76**

Table 1: The seven object types, labeled by the experts. For each type, the amount (noted as "#") of expert labels and gold labels is given together with the detection performance (Recall, precision and F1-rate) of the automated system.

The performance measures varied across the transect. Precision (0.20 to 1.00), recall (0.44 to 1.00) and F1-rate (0.30 to 0.97) showed all no significant correlation with the image area (precision vs. area: 0.25, recall vs. area: 0.15, F1-rate vs. area: 0.26). Figure 2 shows the performances and area across the transect, figure 3 a scatterplot of the F1-rate vs. area. The correlation between F1-rate and area for the object specific analysis are also small and range between -0.12 (Bathycrinus stalk) to 0.38 (Bathycrinus carpenterii).

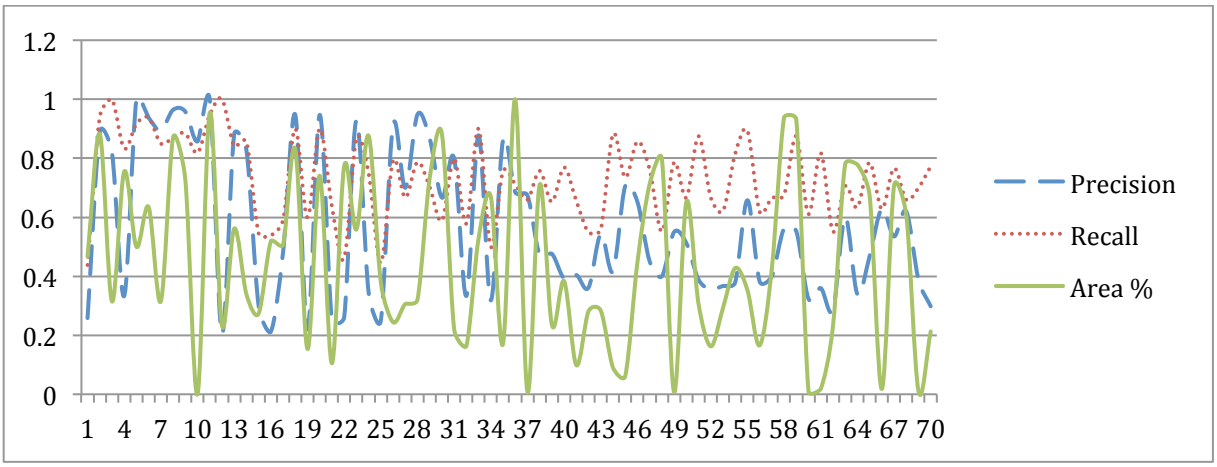

Figure 2: Precision, recall and image area for the 70 images within the transect. The area values were scaled by the minimum and maximum values to lie within the interval [0..1]. The curves show no significant correlation (precision vs. area: 0.25, recall vs. area: 0.15, F1-rate vs. area: 0.26).

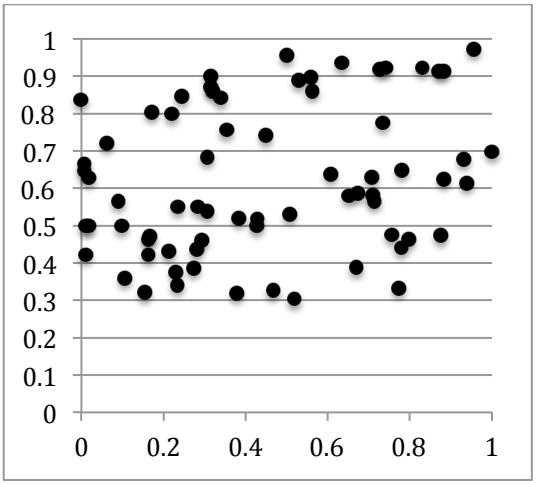

Figure 3: Scatterplot of the per image F1-rate (y-axis) vs. image area (x-axis). The area values are scaled to the interval [0..1].

While most of the label attributes showed no conspicuous relations to each other (e.g. the detection performance is independent of the distance bin), some further hidden parameters were found, two of which will be described here.

The label density per bin showed to be dependent of the bins distance to the lightness peak of the image (see figure 4). The highest label density was gathered in bins 3-6, where bins close to the image lightness peak (1,2) contained less labels and bins the farthest away (7-9) the least amount of labels per area. This shows, that the expert's annotations were influenced by the position of an object within the image.

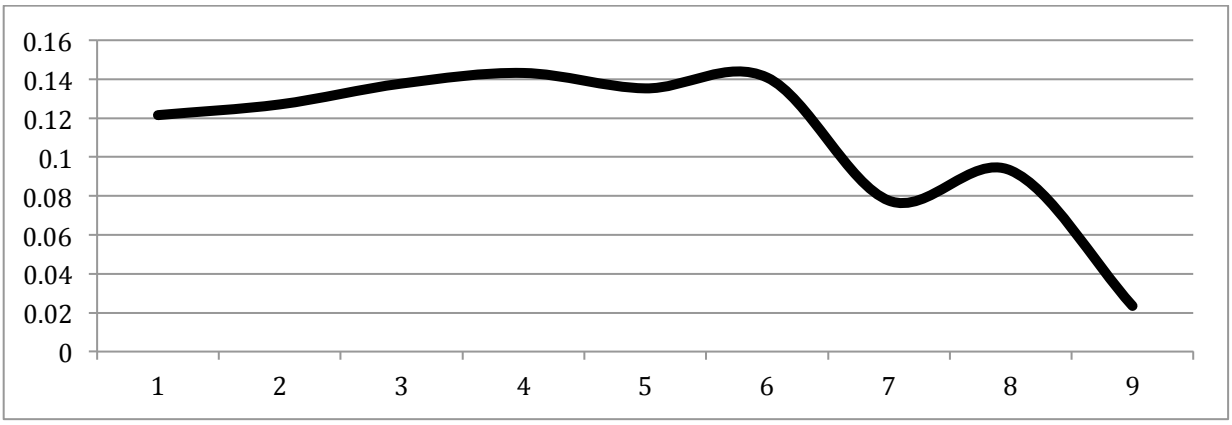

Figure 4: Percentage of label-density per distance bin. The density values (y-axis) were normalized according to the image specific bin areas. The peak of the curve is not at bin 1 where the lightness peak of the image resides but some bins further away, where mediocre lighting conditions occur. The lowest label density is found in bins the farthest away from the lightness peak.

The most significant correlation between label attributes was found for the amount of supporters vs. recall (see figure 5). The recall linearly increased for higher supporter counts. This shows that the system does mimic human experts in its detection ability. Items that are clearly identified by all 5 experts are most likely to be identified by the system as well.

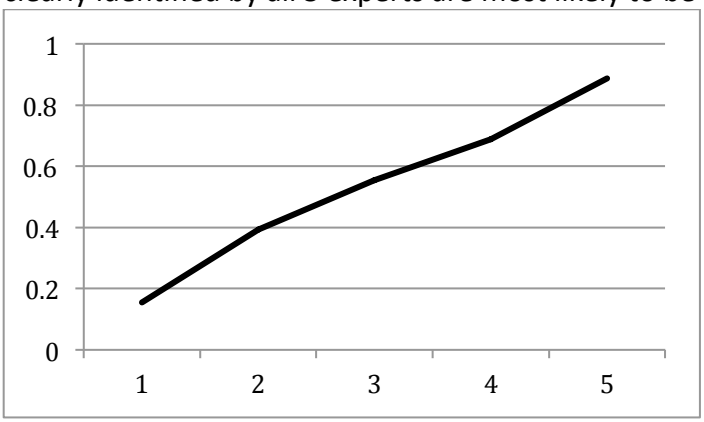

Figure 5: Recall vs. supporter count. Items with a higher supporter count are more likely to be detected by the automated system. Items with 1 and 2 supporters where thought of as untrustworthy and not used for the system parameter optimization. Nonetheless, 20% (s = 1) and 40% (s = 2) of these items are discovered by iSIS.

Discussion:

To solve the problems arising from the artificial light source and camera parameters, we developed an illumination correction algorithm that copes with the differences induced by the distance of the OFOS to the seafloor. The resulting performance values by means of the F1-rate show, that the image area and thus the lighting condition does not influence the detection performance of the system.

The position of objects within the image (regarding the lightness peak position) was identified as a further hidden parameter, in this case influencing the subsequent annotation process. The human experts labeled the original images, not the pre-processed ones. The pre-processing parameters were tuned according to the experts' annotations and hence those images were not available yet. While the innermost distance bins (1-6) feature a comparable label density, a small drop towards the innermost bins was observed. This may be induced by an overexposure of those regions, making bright objects (like sponges and

white anemones) less distinct from the sediment. The lower annotation densities of the outer regions (bins 7-9), may be explained by the blurriness of the image data towards the image corners, making objects difficult to perceive. Additionally, the generally darker regions conceal dark objects (stalks, purple anemones) and especially the shadows of all protruding objects, which seems to be of discriminant importance. The position of an annotation within the image can thus be thought of as a hidden parameter that regulates the annotation process.

The existence of this parameter implies different actions. One solution is to label and classify only central regions of the images. Therefore, some available data (the outer image regions) could not, or not trustworthily be analyzed. Another solution, which requires further expert efforts, could be created by validating the labels in the then available illumination-corrected images.

Validating the labels is also reasonable in making use of the relationship between recall and supporter count. Items that featured the highest values of s=5 were most reliably detected by iSIS. Therefore it should be desirable to gather as many highly trustworthy annotations as possible.

If a re-evaluation is too time-consuming, it might be advisable to include samples with low supporter count (s<3) into the training. Currently iSIS only detects some of these (s=1: 20%, s=2: 40%) and might be able to retrieve more of them, eventually on the cost of lower recall values for the labels with higher supporter count.

As the precision is currently worse than the recall, a third re-evaluation could be put in place for the FPs. We found, that several (about 26 – 35%) of the FPs are caused by objects, that were not labeled at all (s=0). Detailed investigation of those adhered evaluations and the experts' performances will be published elsewhere.

Conclusion:

The consideration of hidden parameters has shown to be important in underwater image analysis. While some effects can be anticipated and be dealt with, others may not be obvious but nonetheless influence subsequent steps or the final detection performance.

Particular challenges arise in the utilization of the essential human expert input. Their knowledge has to be gathered most efficiently and effectively to make sure, that their annotations are trustworthy without incriminating them.

Acknowledgements:

We thank the five human experts for their efforts in annotating the images: Melanie Bergmann, Jennifer Dannheim, Julian Gutt, Autun Purser, James Taylor. We thank Antje Boetius for financial support by the DFG Leibniz program.

References:

[1] Culverhouse P., Williams R., Reguera B., Herry V., Gonzlez-Gil S. 2003 *Do experts make mistakes? A comparison of human and machine identification of dinoflagellates.* Marine Ecology Progress Series 247: 17–25.

[2] Purser A., Bergmann M., Lundälv T., Ontrup J., Nattkemper TW. 2009 *Use of machine-learning algorithms for the automated detection of cold-water coral habitats - a pilot study.* Marine Ecology Progress Series (MEPS).

[3] Rigby P., Pizarro O., Williams S. 2010 *Toward adaptive benthic habitat mapping using gaussian process classification.* Journal of Field Robotics 27: 741–758.

[4] York A., Gallager S., Taylor R., Vine N., Lerner S. 2008 *Using a towed optical habitat mapping system to monitor the invasive tunicate species Didemnum sp. along the northeast continental shelf.* In: OCEANS 2008. pp 1–9. doi: 10.1109/ OCEANS.2008.5152001.

[5] Soltwedel T., Bauerfeind E., Bergmann M., Budaeva N., Hoste E., et al. 2005 *HAUSGARTEN: multidisciplinary investigations at a deep-sea, long-term observatory in the Arctic Ocean.* Oceanography 18: 46–61.

[6] Ontrup J., Ehnert N., Bergmann M., Nattkemper T. 2009 *BIIGLE - Web 2.0 enabled labelling and exploring of images from the Arctic deep-sea observatory HAUSGARTEN.* In: OCEANS 2009 - EUROPE. pp 1–7. doi: 10.1109/ OCEANSE.2009.5278332.

[7] Schoening, T., Bergmann, M., Ontrup, J., Taylor, J., Dannheim, J., Gutt, J., Purser, A., Nattkemper, T.W. 2012 *Semi-automated image analysis for the assessment of megafaunal densities at the Artic deep-sea observatory HAUSGARTEN,* PLoS One

[8] MacLeod N., Benfield M., Culverhouse P. 2010 *Time to automate identification*. Nature 467: 154–155.

[9] Vapnik V. 2000 *The nature of statistical learning theory. Statistics for engineering and information science.* Springer