Ultra-fast segmentation and quantification of poly-metallic nodule coverage in high-resolution digital images

T. Schoening[1], B.Steinbrink[2], D.Brün[2], T.Kuhn[3], T.W.Nattkemper[1*]

[1] Biodata Mining Group, Faculty of Technology, Bielefeld University, PO Box 100131, D-33501 Bielefeld, Germany
[2] saltation GmbH & Co. KG, Bielefeld, Germany
[3] Bundesanstalt für Geowissenschaften und Rohstoffe (BGR), Stilleweg 2, D-30655 Hannover, Germany
*main author: Tim W. Nattkemper, tim.nattkemper@uni-bielefeld.de

ABSTRACT

The exploration of the seafloor regarding the abundances of poly-metallic nodules has received growing attention recently. Image-based exploration using camera-equipped deep-sea observation systems has been introduced successfully to increase spatial resolution in the exploration process but the evaluation and interpretation of the huge amount of image data creates a serious bottleneck. In this paper we present a new software solution to the problem of automatic detection and segmentation of poly-metallic nodules in underwater images, which is not only accurate but so fast that real time image analysis is definitely in reach. Thus, an application on an AUV (Autonomous Underwater Vehicle) seems possible in the near future.

Keywords: poly-metallic nodules, underwater image analysis, real time image processing, AUV

Introduction

The economic development of the last two decades has led to a rapid development of the manufacturing industries in countries with emerging economies. This has a considerable impact on the price development for resources such as oil, gas, or metals. The oceans, with the majority of their area still unexplored, are catching more and more attention in this context, adding the components of mineral / metal resources to their classic economic aspects like fishery, oil/gas or offshore wind parks. Here we will consider the issue of exploration of underwater seafloor areas which are covered with poly–metallic nodules. In principle, those are observed all over the world, but hotspots have been reported in the Indian and as well as in the Pacific Ocean. To enable a development of technological strategies for harvesting these

nodule fields, which are effective and efficient, a quantification of the topological nodule distribution and the expected masses of the nodules is of fundamental importance.

This paper focuses on work within the German license area for the exploration of Mn nodules in the eastern equatorial Pacific (Kuhn et al.; 2011). Sonar backscatter has been used successfully in the German license area to distinguish seafloor areas covered by Mn nodules from areas devoid of them (Kuhn et al. 2011). However, the significance of this reported survey is limited due to the low resolution of the sonar data of ~120mx 120m per pixel. As a consequence, this source has been extended by an additional one for ground truthing: digital images / video. Camera transects in the same area were captured using a towed on-line camera/photo system which enables a high resolution visual assessment of the seabed with an average spatial resolution of 0.5mm^2 / pixel. However, the high resolution comes with the serious drawback of a bottleneck in data analysis. An exploration of the entire German area would create approximately n*10^2 Petabyte of image data, making an exhaustive manual analysis impossible. This limits the applicability of imaging to small selected sub-regions where it can be correlated to the sonar backscatter data. Since underwater image analysis is gaining more attention in the last years in the context of habitat mapping and taxonomic classification (Williams et al 2012, Ontrup et al. 2009, Purser et al. 2009, Schoening et al. 2012A) we recently proposed to apply machine learning – based image analysis for the tile-wise quantification of poly-metallic nodule coverage in categories of 0%, 20%, 40%, 60%, 80%, and 100% (Schoening et al, 2012b).

Despite the positive initial results, we identified two pressing problems that come with image – based quantification of poly–metallic nodules. First, since the nodules are analyzed in images, the computational classification was restricted to counting all the pixels in one tile that have color features like those observed for nodules. So the methods made (almost) no difference between a pattern of a small number of huge nodules or a huge number of small nodules. But given the background knowledge about the non-isotropic growth pattern of nodule volumes, the resulting masses for these two scenarios should be expected to be significantly different from each other. Thus, we need to consider the size and shape of the polymetallic nodules so we have to leave the tile-wise quantification concept and need a full segmentation that gives the full information about size and shape for each nodule. Second,

the processing time for one transect was between 10-12 hours on a standard PC that can be used to analyze the images off shore on the ship. This time needs to be shortened significantly so the user can react to the results as soon as possible.

In this paper we present a new algorithmic solution to the problem of poly-metallic nodule segmentation that is able to solve the two problems listed above. The nodules are segmented and mapped to user-defined size categories, which allows a more detailed correlation to the backscatter data and represents an important step towards mass quantification in images in the future. And due to algorithmic improvements and making use of the PC graphics card we were able to shorten the processing time from 52 secs / image to less than 0.5 secs / image when processing a single image and an average of 0.18 secs / image in batch processing. This enormous speed up even makes the application on AUV (Autonomous Underwater Vehicles) a reasonable option, which paves the way to completely new strategies for initial exploration, i.e. maybe imaging could be used as a mapping technique and not "just" for ground truthing.

Background and Material

In this paper we present results obtained for two different image transects (833 and 1267 images each) recorded with a towed camera-sledge (OFOS) in the Clarion / Clipperton area of the equatorial Pacific which was explored under the German license in 2010 (Rühlemann et.al. 2010). The original color images (4224 x 2376 pixels, 8bit/color channel) display an area of 1.3 to 5.3 m^2 depending on the varying distance to the ground, thus showing changing illumination characteristics regarding the color and illumination cone. To exclude some of the effects of non-homogenous illumination and a slight fish-eye effect on the image border, we focused our analysis and the expert exploration on the center part of the images with a size of 2000 x 1000 pixels. All these images were subject to the nodule segmentation algorithms presented in this manuscript and for all the images sonar backscatter results were available for a following correlation analysis (see Results section).

Methods

To achieve an image segmentation, each pixel *p* in the image must be labeled to be either part of a poly-metallic nodule or not. So the segmentation can be considered as an application of a labeling function $L(p) = \{0,1\}$ that assigns each pixel to the classes "poly-metallic nodule" ($L(p)=1$) or "no poly-metallic nodule" ($L(p)=0$). The mapping is achieved in three steps (1.) pre-processing, (2.) H^{2}SOM mapping and (3.) segmentation that are briefly described below (see also Figure 1). To achieve substantial speed up several optimizations have been applied described in the last subsection.

Preprocessing

Each image was subject to two pre-processing routines. First the illumination cone was removed using an illumination cone model of a Gaussian function. Second, the color histograms was shifted so that the peak of the illumination histogram was at 127. Both processes run fully automatic (see details described in Schoening et al. 2012a), however the first step needs a lot of computing time.

H^{2}SOM mapping

To map each pixel $\boldsymbol{p}$ to a label $L(\boldsymbol{p})=\{0,1\}$, the pixel is first represented to a feature vector $\mathbf{x}(\boldsymbol{p})$ that describes the color features in a 7 x 7 neighborhood around this pixel. To this end, the histogram of this neighborhood in each color channel (R=red, G=green, B=blue) is reduced from 256 to just 16 bins. The frequencies of the bins in all three histograms are concatenated to a 48-dimensional feature vector (see also Schoening et al, 2012b). Next, the feature vector $\mathbf{x}(\boldsymbol{p})$ is mapped to one of 160 clusters. The clusters are trained using a H^{2}SOM (Purser et al. 2009, Schoening et al, 2012b) in a prior step using a small subset of 0.5% of the transect data. Each cluster consists of a small set of feature vectors, which are alike (i.e. show very similar color histogram features) and is represented by one so called prototype vector $\mathbf{u}^{(i)}$ (i=1,…,160). Since the clusters are trained using the H^{2}SOM algorithm, each cluster can be assigned to a cluster color $\mathbf{c}^{(i)}$ = (R,G,B), so two clusters which are very similar get very similar clusters (which is referred to as *topology preservation* in the machine learning com-

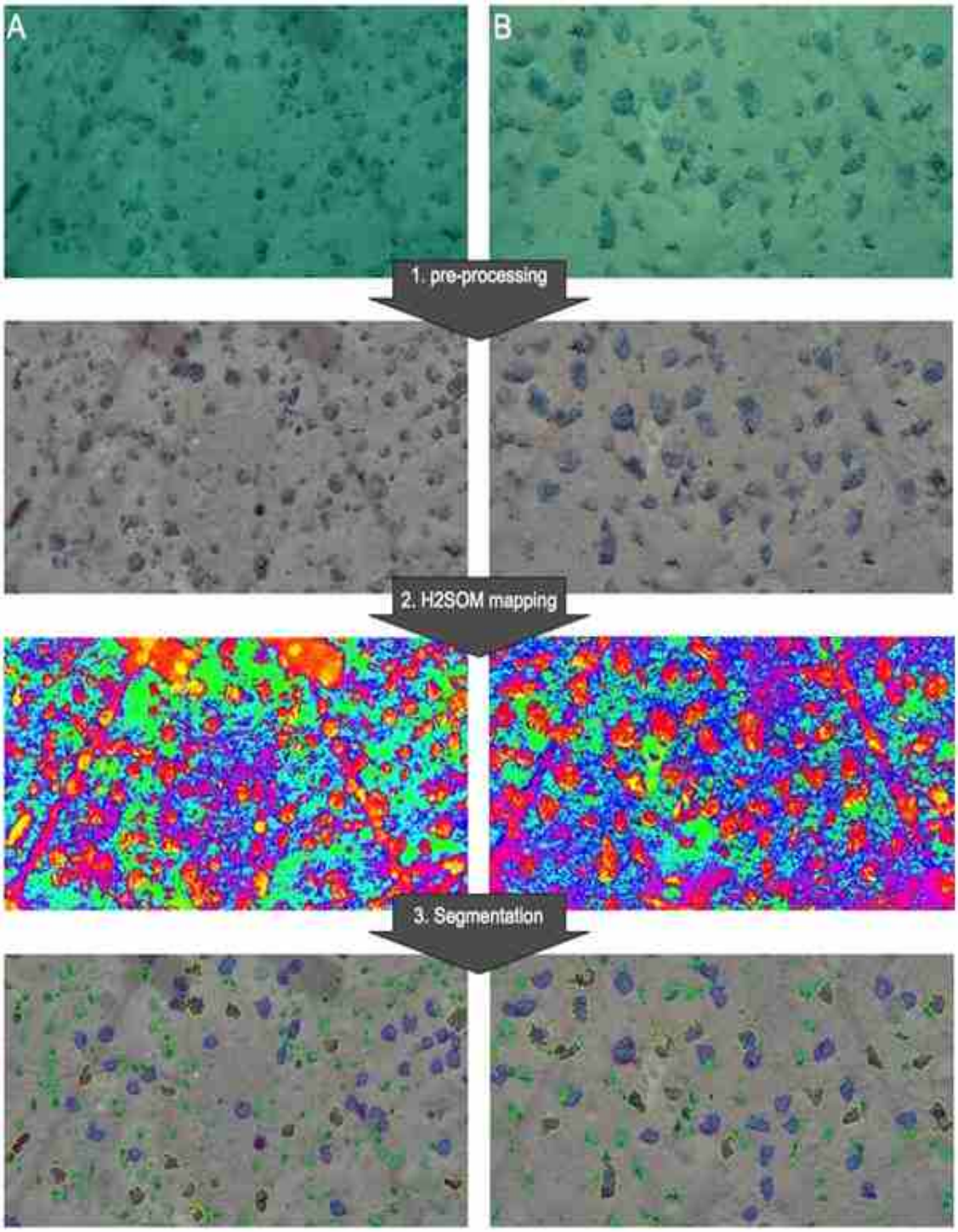

Figure 1: : For two example images A and B from the transects the full segmentation process is illustrated with intermediate results. First, the input images are pre-processed to correct illumination artifacts caused by varying distance to the ground (1. pre-processing). In a next step, all pixel features are mapped to clusters in a H^2SOM, which is visualized with pseudo color images. Each color encodes a cluster index. Similar colors are assigned to similar clusters (2. H^2SOM mapping). The cluster map is post-processed to identify separated regions of single poly-metallic nodules, which are outlined in the last row (3. segmentation). Here colors encode the size interval of the nodules as chosen by the user: green: 0.25 - 3.14 cm^2, blue: 3.14 - 12.57 cm^2, yellow: 12.57 - 28.27 cm^2 pink: >28.27 cm^2.

munity). The mapping of feature vectors to clusters is visualized in Figure 1 (third row). Each pixel $\boldsymbol{p}$ was first represented with its feature vector $\mathbf{x}(\boldsymbol{p})$, which has been assigned to that cluster $\mathbf{u}^{(i)}$, that has the most similar feature composition (usually referred to as BMU (Best Matching Unit)). The mapping result can be visualized as a pseudo color map (e.g. the third row in Figure 1), where each pixel is drawn with the color $\mathbf{c}^{(i)}$ of its individual BMU. So if two pixels have the same color, they have been assigned to the same cluster $\mathbf{u}^{(i)}$. In Figure 1 one can see, that the nodules are mapped to colors ranging from yellow via orange to red.

Since the 160 clusters have been visually assessed by a human expert, who classified each cluster as being a cluster of feature vectors from poly-metallic nodules or not, this information can be used to implement a classification function $C(\mathbf{u}^{(i)}) = \{0,1\}$ (0=no poly-metallic nodule, 1=poly-metallic nodule). The final segmentation $L(\boldsymbol{p})=\{0,1\}$ is obtained by mapping each pixel to 1 if the best matching prototype to this pixel (i.e. its feature vector) is classified as a poly-metallic nodule cluster (i.e. $C(\mathbf{u}^{(i)}) =1$) and 0 otherwise. This way, a binary image is obtained.

Postprocessing

The binary output images are post–processed for the final result. In a first step, morphological operators (i.e. an *opening* operation) are applied to delete small regions of just a few pixels and to close gaps in the regions (Gonzalez and Woods, 2001). In the second step, a list of regions (poly–metallic nodules) are determined with size in pixels and cm^2. The latter can only be calculated from the region sizes given in pixels if the spatial pixel resolution is available. In the presented case, laser pointer marks were used for this purpose. To compute the finals result, the user defines areal intervals for M nodule area size classes and the abundancies for each size class are computed (in the presented case C_1: 0.25 - 3.14 cm^2, C_2: 3.14 - 12.57 cm^2, C_3: 12.57 - 28.27 cm^2 C_4: >28.27 cm^2, so M=4, see last row in Figure 1).

Towards real time segmentation

The original implementation consisted of a single sequential pipeline, processing one image at a time. Table 1 shows a breakdown of the times each processing stage takes. The major bottlenecks in this initial version were the H^2SOM mapping and the assignment of pixels

identified as poly-metallic nodules to individual nodules taking about 18 and 28 seconds respectively. Another time-intensive operation was the illumination correction in the pre-processing (Gaussian filter and color histogram transformation) which took another 5.5 seconds.

Pre-processing speed-up: The pre-processing lends itself well to parallelization and therefore GPU (Graphics Processing Unit) processing. Since we were already using OpenCV, which supports CUDA-based (CUDA Programming Guide, 2011, S. Cook, 2012) GPU processing for a lot of operations, using CUDA was a natural choice. Using the GPU convolution function provided by OpenCV (G. Bradski and A. Kaehler, 2012) to perform the Gaussian blur reduced the required time from 2.8s to just 80ms per image. A new efficiency oriented implementation of the histogram transformation in CUDA reduced the time for this operation from 2.8s to just 7ms.

H^2SOM mapping and post-processing speed up: The H^2SOM mapping of one pixel consists of a) the construction of its feature vector and b) the search for the best matching unit. Since these two operations can be carried out for different pixels independent from each other, a) and b) are subject to a parallelization as well. To this end the image is split into tiles which are processed in parallel by a number of threads. To limit the time expensive access to the GPU we apply the "shared memory feature" so information from overlapping pixel neighborhoods can be used efficiently. In addition, CUDA's "constant memory" function is applied to increase the efficiency in the BMU search, when all threads look up identical prototype vectors. Another speed up was achieved by neglecting those pixels, which were cut of in the post processing anyway.

Finally, the classification step could be slightly accelerated by switching the processing order from column-wise to row-wise which matches the layout the image has in memory and is therefore more cache-friendly. The most expensive part of the post-processing, the identification of regions, could be accelerated as well based on the constraint that regions should be non-overlapping. This allows to draw a special 32bit index image which can be transformed into the final label image very fast.

Stage	Operation	Runtime before optimizations	Runtime after optimizations	Using CUDA	Change
Preprocessing	Image loading	150ms	150ms	N	0%
	Gaussian Filter	2800ms	80ms	Y	-97%
	Histogram transformation	2770ms	7ms	Y	-99%
H^2SOM mapping	H^2SOM mapping	18100ms	90ms	Y	-99%
Postprocessing	Downsampling	40ms	–	–	
	Classification	20ms	7ms	N	-65%
	Opening	2ms	2ms	N	0%
	Gap removal	180ms	24ms	N	-87%
	Region building	28250ms	90ms	N	-99%
	Analyze	24ms	24ms	N	0%
	Total	52336ms	474ms		-99%

Figure 2: Execution times for a single image for the individual processing steps, before and after optimizations have been applied. All times were measured on an i7-3770 running at 3.5GHz with a GeForce GTX 670 running Debian Linux using GCC 4.7.3, OpenCV 2.4.5 and CUDA SDK 5.0.

Parallelization optimization: Experiments indicated that the original sequential implementation makes poor use of the available resources (CPU and GPU), introducing idle times for the GPU as well as the CPU. So for a further improvement, we organized the entire process into the three tracks image loading (1), GPU processing (2) and post-processing (3) and assigning these to three threads, which communicated via work queues.

Future improvements towards real time processing: The primary remaining bottleneck in the presented set up is the GPU. As the H^2SOM mapping on the GPU seems to be limited by the number of available registers on our CUDA compute capability 3.0 hardware, a next step could be introducing a CUDA compute capability 3.5 card which offers 255 instead of 63 registers. Additionally, the code could be extended to support multi-GPU configurations.

Results

The segmentation results were compared to those obtained with sonar backscatter signals. To this end, the output of the segmentation was plotted over time together with the backscatter signal. The segmentation result fitted well to the backscatter signal, if the H^2SOM training and the segmentation was applied to the same image material (as evaluated by one of the

authors, T.K.). If the H^2SOM was trained on one transect A and applied to a transect B afterwards, the results were spoiled due to the fact, that the average distance between the camera and the ground varied between the two data sets of A and B. However, the training of an H^2SOM takes ~50 minutes, so an adaptation step to changing distance is possible. Nevertheless, one has to re-consider the assignment of prototypes to labels/categories $C(\mathbf{u}^{(i)})$ = {0,1} (0=no poly-metallic nodule, 1=poly-metallic nodule) so time for a visual assessment of each prototype must be included in the calculation as well. In addition, if the distance to the ground is recorded, the images can be transformed to have constant pixel resolution which will be subject to future developments.

Discussion

We have presented a new algorithmic approach to the problem of poly-metallic nodule segmentation in underwater image data. The results showed sufficient accuracy and the substantial acceleration in processing time presented in this paper motivates developments towards real time image pressing. This would allow an implementation and application even on AUVs (Autonomous Underwater Vehicles) which would pave the way to new strategies for early exploration of potential nodule fields. Of course this would be condition to an extension that would reduce the effects of changing ground distances as discussed above. In addition, the training data needs to be extended to reflect changes in sediment characteristics as well. Another interesting aspect is, that we are now able to achieve a full segmentation. The shape of the individual nodule regions could be used for a more "intelligent" estimation of masses per regions. Since large parts of the nodules are covered, a direct mapping from pixels to cm^2 to mass must lead to non-optimal results. So new strategies and concepts for the mass estimation are required. We believe that the segmentation results with our algorithms provide a new and promising basis for such an improvement.

Acknowledgements: This work was supported by the Bundesanstalt für Geowissenschaften und Rohstoffe (BGR), D-30655 Hannover, Germany, and the BMWi grant 03SX344A „TIMM – IS2U".

References

Kuhn, T., Rühlemann, C., Wiedicke-Hombach, M. (2011). Development of Methods and Equipment for the Exploration of Manganese Nodules in the German License Area in the Central Equatorial Pacific, Proc. Ninth (2011) ISOPE Ocean Mining.

Gonzalez, R.C. and Woods, R.E. (2001). Digital Image Processing, 2nd Edition. Prentice Hall International, ISBN: 0201180758.

Ontrup, J., Ehnert, N., Bergmann, M., Nattkemper, T.W. (2009). Biigle - Web 2.0 enabled labelling and exploring of images from the Arctic deep-sea observatory HAUSGARTEN. IEEE Oceans 2009, pages 1-7.

Purser, A., Bergmann, M., Lundälv, T., Nattkemper, T.W., Ontrup, J. (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 (2009) doi: 10.3354 / meps08154

Rühlemann, C. and Shipboard Scientific Party (2010). *Cruise Report SO-205 MANGAN*. Bundesanstalt für Geowissenschaften und Rohstoffe, Hannover, 112 pp.

Schoening T., Bergmann M., Ontrup J., Taylor J., Dannheim J. (2012a). Semi-Automated Image Analysis for the Assessment of Megafaunal Densities at the Arctic Deep-Sea Observatory HAUSGARTEN. PLoS ONE 7(6): e38179. doi:10.1371/journal.pone.0038179.

Schoening, T., Kuhn, T., Nattkemper, T.W. (2012b). Estimation of poly-metallic nodule coverage in benthic images. The 41st Conference of the Underwater Mining Institute.. October 15–20, Shanghai, China.
Williams, S. B., Pizarro, O., Jakuba, M. V., Johnson, C. R., Barrett, N., Babcock, R. C., Kendrick, G. A., Steinberg, P. D., Heyward, A. J., Doherty, P. J., Mahon, I., Johnson-Roberson, M., Steinberg, D., Friedman. A. (2012). Monitoring of Benthic Reference Sites Using an Autonomous Underwater Vehicle. ZEITSCHRIFT FEHLT!!! 19(1) 73-84.

Cook, S. (2011). Cuda Programming: A Developer's Guide to Parallel Computing With Gpus (Applications of Gpu Computing Series), S. Cook, M. Kaufmann Publishers.

CUDA Programming Guide, C NVIDIA 2011,
http://www.nvidia.com/object/cuda_home_new.html

Bradski, G., Kaehler, A. (2012). Learning OpenCV: Computer Vision with the OpenCV Library. O'Reilly Media http://opencv.org/.

Tim W. Nattkemper and his Biodata Mining Group develop new methods for the automatic analysis of underwater images since 2007. The group combines classic image processing routines with special machine learning architectures to design solutions for quantification and classification in underwater image and video data which are both: effective and efficient. TWN collaborates successfully with partners from academia/public funding (AWI (Bremerhaven, Germany), BGR (Hannover, Germany) as well as with industry (STATOIL ASA). In 2012, TWN started a collaboration with a professional software company, Saltation GmbH (Bielefeld, Germany), to transform the scientific software solution to market-ready products that can be applied in many fields such as underwater mining, marine biology, environmental monitoring.