Rapid image processing and classification in underwater exploration using advanced high performance computing

Timm Schoening
Deep Sea Monitoring
GEOMAR Helmholtz Centre for Ocean Research
Germany

Daniel Langenkämper
Biodata Mining Group
Bielefeld University
Germany

Björn Steinbrink
saltation Marine IT
Germany

Daniel Brün
saltation Marine IT
Germany

Tim W. Nattkemper
Biodata Mining Group
Bielefeld University
Germany
tim.nattkemper@uni-bielefeld.de

***Abstract*—Computational underwater image analysis is developing into a mature field of research, with an increasing number of companies, academic groups and researchers showing interest in it. While on the one hand, the basic question is addressed by many groups, how algorithms can be applied to automatically detect and classify objects of interest (OOI) in underwater image footage, on the other hand the questions for efficiency and performance, i.e. the time a computer (or a compute cluster) needs to perform this task, has received much attention yet. In this paper we will show, how nowadays methods for high performance computing like parallelization and GPU computing via CUDA (Compute Unified Device Architecture) can be used to achieve both, image enhancement and segmentation in less than 0.2 sec per image (4224×2376 pixel) on average, which paves the way to real time online applications.**

Keywords: underwater image analysis, marine informatics, underwater image informatics, segmentation, GPU, parallelization, image enhancement

I. Introduction

To explore and monitor underwater habitats and resources, imaging is applied using different devices carrying high resolution digital cameras such as AUV (Autonomous Underwater Vehicle), ROV (Remotely Operated Vehicle), OFOS (Ocean Floor Observation System) or even benthic landers / stationary observatories [1], [2], [3]. Until now, only a small subset of the image / video data has been analyzed and the majority of this data have been anaysed manually by direct visual inspection on a computer screen. Experts of their disciplines, be it taxonomist, ecologist, geologist or else, browse the images in an annotation software and mark points or regions as objects of interest (OOI), associated to some semantics. Various software products have been developed for the annotation of video frames [4], random points [5] and manually chosen points of interest or regions of interest which are associated with a semantic category (like from biological taxonomy, morphotypes or some pre-defined technical ontology). Prominent examples for such software systems are single work space desktop applications like NICAMS (NIWA Image Capture and Management System), VARS (Video Annotation and Reference System) und SQUIDLE or web tools that support collaboration such as BIIGLE (Benthic Indexing, Graphical Labeling and Exploration)[6], CATAMI (Collaborative and Automated Tools for Analysis of Marine Imagery)[1] or MAMAS[2] [7].

Despite these efforts to support an efficient and effective manual data analysis by human domain experts, it is evident that the rapidly growing piles of digital marine image data cannot be analyzed by human observers alone. One serious problem in manual annotation is caused by the general cognitive limitations of humans regarding visual assessment and semantic annotation, which has been a subject of study since the 1950s in the context of radar signal interpretation, PEP smear analysis and recently in marine image interpretation as well [8], [9], [10]. One fundamental problem here is, that in almost all studies and projects there is no clear idea or definition of an the *quality* of an annotation result, which is usually described by the *accuracy* and the *precision* of an annotator.

Another problem is the sheer size of the image data archives. Depending on the habitat's biodiversity, the species densities, the pixel-to-centimetre ratio, the conspicuity and abundance of other objects of interest, annotation times per image can vary significantly, but can easily reach more than one hour per image [11], [9].

In cases where the whole image is to be segmented into objects of interest and the sediment background, this process can take even longer than two hours, especially when hundreds of OOI are apparent. An example of such a challenging task is the segmentation of poly-metallic nodules that occur in various areas on the seafloor. These nodules are a prospective resource and are currently considered for deep sea mining (again) as specific resource prices rise. Assessing the nodule abundance and density in an explored area requires visual ground truth data acquired rom image transects, that can be correlated with large-scale habitat maps from side scan or multi beam data [12]. This is a common method in marine geology and resource exploration and also applies to other marine resource assessments [13], [14]. Apart from ecological conservation aspects that currently prevent the extraction of marine resources in international waters [15], [16], [17], [18], an accurate assessment of resource abundance and density is also required to optimize the mining process, i. e. to make the extraction cost effective. In this specific case of marine image analysis, initial approaches have been published [19],

[1]http://catami.org/
[2]http://www.saltation-marine.com/

978-0-933957-43-5 ©2015 MTS

[20] but generally applicable solutions are still missing due to variations in the image qualities caused by changes in acquisition gear (i. e. cameras, lighting, platforms) between different exploration cruises. As a consequence, image analysis algorithms need to be adapted to new image footage which can be a time consuming and computational expensive. Thus, the development of algorithmic solutions needs to focus more on performance and efficiency in this context.

Computational image analysis of marine data is in general gaining popularity lately, because of the intially stated problems [11], [21]. The scale of the task is so large, that even the application of standard desktop software solutions can fail. And this bottleneck in underwater image analysis gets more serious due to the continuous increase in spatial and temporal resolution. So although some first works have been published that show clear potential for computational underwater image analysis the question of computational performance and speed must also be addressed now. This question must be considered in front of the following background issues: First, expedition time on the ship is expensive, so computational analysis on board during the cruise would be beneficial since it would give biologists / geologists a quick insight into the the data content. One option would be to install large scale compute infrastructure on board the research vessel, like for instance extending the ship's infrastructure with a compute cluster or by developing a portable system that can be deployed if needed. Such a cluster infrastructure would require several dozen CPUs and accompanying hardware, amounting to costs of several ten thousand euros. These investments limit the usefulness of the cluster approach. Thus, it is important to process the images as fast as possible with the compute power on board (i.e. standard PCs) since employing external (i. e. on-shore) compute power for large data collections during a cruise is also rarely possible nowadays.

The second background issue is given in the context of monitoring and autonomous systems such as AUV or crawlers [22]. Such systems could benefit from basic pattern recognition models that run online embedded in the system and in real time to enhance image quality or to detect events or objects that can be taken into account in navigation and signal recording.

In this paper we will show, how algorithms for image enhancement and classification can be re-designed to reach computation times that make an application of these algorithms in the aforementioned scenarios reasonable.

II. Materials

We demonstrate the progress in computation acceleration with image data from underwater mining exploration acquired by the BGR (Bundesanstalt für Geowissenschaften und Rohstoffe (engl.: Federal Institute for Geosciences and Natural Resources in Germany) in the Clarion Clipperton Fracture Zone (CCZ) in the central eastern Pacific. The images were acquired with a downward looking camera, mounted on a towed camera platform and show the ground / sediment, (partly occluded) poly-metallic nodules as well as megafauna. Laser pointers provided scaling information and automated laser point detection was used to calculate the real world footprint [23]. On average, the images footprint is 1-5 square meters and each images has a resolution of 12MP. Figure 1 shows an example image that was acquired in 2010 during cruise SO205 of the German research vessel FS Sonne.

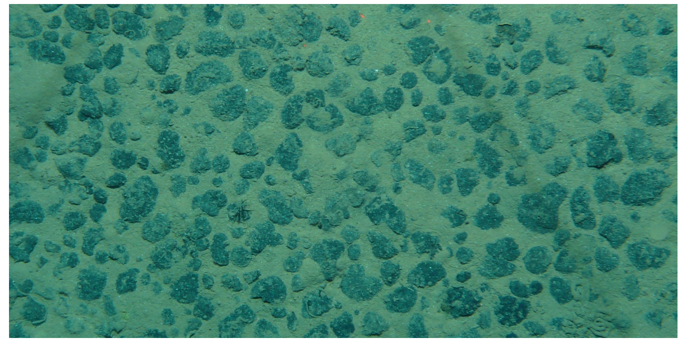

Fig. 1. One example image showing the challenges of automated nodule segmentation: color variances within an image due to the illumination drop-off towards the corners, sediment coverage of nodules and adjacent nodules that have to be split up to estimate nodules abundances correctly. Other problems could be bioturbations of the sediment as well as epifauna, growing on the nodules.

III. Methods

We will investigate the acceleration potential for three different algorithms: First, the method for image enhancement, referred to as fSpice that applies a computationally tuned model based illumination correction and a color balancing function to compensate attenuation effects [9], [24]. Second, the application of a neural network (referred to as H^2SOM) for poly-metallic nodule segmentation using pixel classification [25], [12]. Third, the adaption tpo new image footage, i. e. the training of the H^2SOM using a subset of the images from a new transect. All three algorithms have been subject to an acceleration optimization on standard desktop hardware (Intel 3rd gen Core i7-3770, 32GB RAM, Nvidia GTX 670):

a) fSpice

The color and illumination correction with fSpice has been recently compared to other pre-processing algorithms and has shown to produce results of superior or equal quality [24]. Its algorithmic definition lends itself well to parallel processing on a GPU (Graphics Processing Unit). The method consists of a Gaussian filtering to model the illumination cone pattern and using the model to compensate thois illumination cone. This step can be efficiently executed on a CUDA[3] (=Compute Unified Device Architecture) enabled hardware. The second part of fSpice is a histogram equalization step that takes the individual image characteristics into account to normalize the images within a transect. The result of fSpice is shown in Figure 2. Both steps lend themselves very well to parallelization on CUDA hardware, thus adding a significant performance boost. To perform the Gaussian filtering on CUDA-enabled hardware, we employed algorithms readily available in the OpenCV library. For the histogram equalisation we implemented custom CUDA kernels to enable parallel execution and avoid expensive data transfers between the host system and the CUDA hardware.

b) H^2SOM segmentation / pixel-classification

Segmentation of the images is achieved by binary pixel labeling and post-processing. The pixel labeling process is based on a vector quantization / prototype clustering of the

[3]NVIDIA Corporation, CUDA Programming Guide 1.0, http://www.nvidia.com,2007

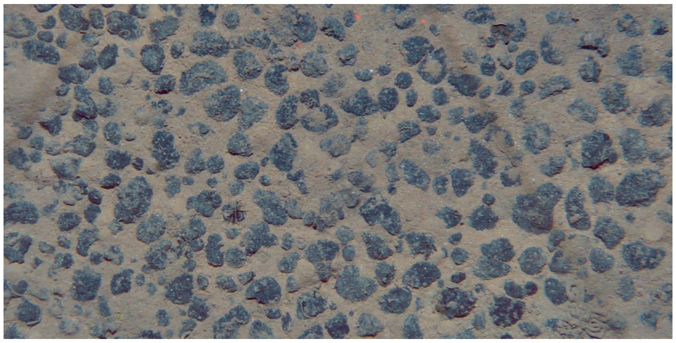

Fig. 2. The result of fSpice application for the image shown in Figure 1. Notice the improved contrast between the sediment background and the nodules as well as the more natural colors that even resemble the natural appearance of the CCZ ground as raised with clamshells/grabbers and inspected out of the water.

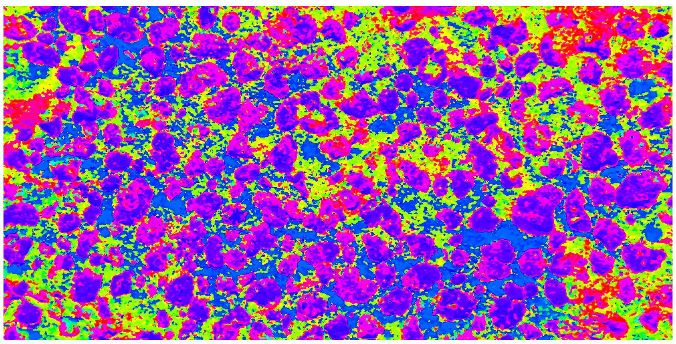

Fig. 3. Visualisation of the H^{2}SOM clustering: The prototype / cluster index are encoded as colors, each cluster is assigned to a different color, dependent on its position in the H^{2}SOM grid. Due to the topology preservation feature of the H^{2}SOM, prototypes with similar features are assigned to similar colors.

visual feature space extracted from a low number of example images. A H^{2}SOM [11], [12] is used for clustering since it supports easy straightforward labeling of prototypes to be background or poly-metallic nodule-like. As a consequence to using a vector quantization-based pixel classification for segmentation, this step can be accelerated significantly by parallelization since image subregions / pixels can be processed independently regarding their relative positions to the learned feature prototypes. For the segmentation, a trained H^{2}SOM, i. e. a quantization of vector space by the H^{2}SOM prototypes is needed, as well as a feature representation of the pixels as H^{2}SOM training input. This feature representation consists of 16 bin histograms of the three RGB color channels so the feature vectors are 48-dimensional. In the segmentation step of one image, the feature vectors of alle pixels are assigned to their best-matching unit (BMU) using the trained H^{2}SOM. That way, every pixel is mapped to the index of the H^{2}SOM prototype that has the lowest Euclidean distance to it in the feature space. One beneficial characteristic of the H^{2}SOM is its ability to projects the feature space onto a two-dimensional embedding, allowing to visualize the clustering in the image. This embedding is distance preserving such that similar colors represent similar features and thus probably similar objects or pixels belonging to the same segment (see Figure 3).

By manually assigning classes (i. e.nodule, sediment) to each of the H^{2}SOM prototypes, the pixel mapping was transformed to a binary image, where 1-pixels (white) represent nodules and 0-pixels (black) represent the sediment background (see Figure 4). Image processing in the form of morphological operations, blob detection and blob splitting through shape analysis was applied to determine the outlines of individual nodules. Fusing these pixel-based nodule statistics (nodule amount and size distribution per image) with the real-world image footprint then allows to estimate these values in centimeters (see Figure 5). Additionally, seafloor coverage in percent can be measured.

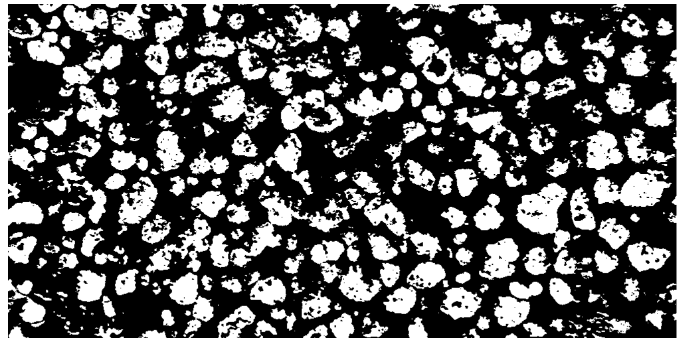

Fig. 4. Binary mapping of each pixel of Figure 1 to either the nodule class (white) or the sediment background class (black).

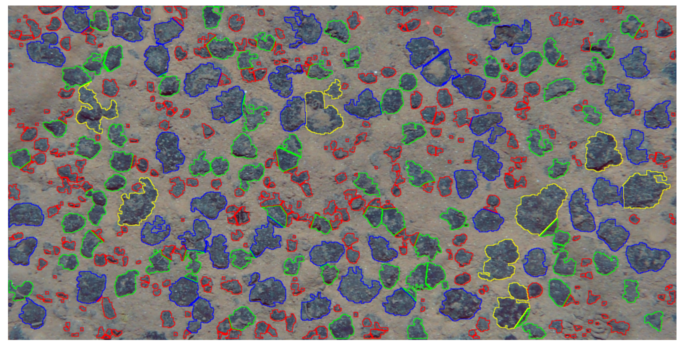

Fig. 5. Final result of the nodule segmentation. Nodule size groups are visualized by colored outlines around the detected segments (red: small, green: middle, blue: large, yellow: very large).

The feature computation and BMU search are both executed on the CUDA hardware which resulted in a tremendous speedup. To make efficient use of the available CUDA hardware, images were split into blocks that are processed independently and are small enough to fit entirely into GPU-registers, thus avoiding expensive memory access. Instead of creating a 48-dimensional dataset of feature vectors first and then mapping this dataset onto the BMUs, these steps are fused together in order to avoid the creation of this intermediate dataset for the whole image. As usual with CUDA, these computations happen in parallel within each block, as well as processing multiple blocks in parallel. The pixel mapping and the final image processing still happens on the CPU, but multiple images are processed in parallel.

It is important to note at this point, that the segmentation step is not carried out for the entire image, since the optics create serious distortions at the image borders. Thus, only a central rectangular subregion of 2112×1188 pixel is subject to the segmentation step.

TABLE I. THE TABLE SHOWS THE SPEED UP FOR THE THREE SCENARIOS. (TODOBJOERN: PLEASE CHECK NUMBERS)

Image processing step	processing time using standard implementation	processing time with new high performance implementation
a) fSpice	5.77 sec	0.087 sec
b) H2SOM segmentation	18.1 sec	0.237 sec
c) H2SOM training	1900 sec	650 sec

c) H2SOM training
The training can hardly be parallelized, but by carefully restructuring and optimizing the code for the given H^2SOM grid structure and tuning the code to account for the features the target hardware platform (see laptop specifications above), a notable speedup could be achieved. For instance, besides providing as much information to the compiler as possible (like making all relevant sizes and variables known static), we used data structures that provide good cache locality and allow for optimal usage of the vectorization units available in modern hardware. Further speedups were achieved by using a fast XorShift random number generator instead of the one supplied by the system.

IV. RESULTS

The resulting computation time for the three computational steps a) pre-processing, b) segmentation, and c) H^2SOM training are shown and compared to the computation time in table I. The time for analyzing one image is the sum of a) and b) and sums up to 0.324 sec per image, but averages to less than 0.2 sec per image in batch processing thanks to parallel processing. This process uses a pre-trained H^2SOM that does not require a large volume of working memory and is implemented on "off the shelf" hardware (see above).
The H^2SOM training would be needed to adapt the system to new data sets (for instance acquired with a new camera or a new illumination system or in a different area with different visual qualities. The new training time around 10 min allows a fast and responsive learning step during the cruise / the ship time and can now be applied instantaneously after acquiring the new image data with an exploration. This enables a fast and effective segmentation during the cruise and an integration of the image analysis results in the decision processes during the exploration cruise.

V. DISCUSSION

Our results show that the computation speed can be increased by several orders of magnitude. The new implementation indicates that the application of sophisticated algorithms such as machine learning or neural networks can be efficient even on standard laptops and on board. Thus, ship time can be used more effectively with data analysis which is currently often needs to be delayed until the end of the cruise. These computation times may also be considered in the context of AUV development in the future. Analyzing AUV footage in real time could allow for on-the-fly adjustments of the AUVs route, thus enabling the vehicle to cover more ground in regions deemed interesting. In the light of the enormous amounts of images that are gathered and stored every day, transferring the software described here into a scalable cloud-infrastructure is also a very promising approach for future developments.

ACKNOWLEDGEMENTS:

This work was funded by the Bundesministerium für Wirtschaft und Energie (BMWI, FKZ 03SX344A). We thank T. Kuhn from BGR for the provision of underwater images and the fruitful discussions within this project. The underwater pictures were taken with a towed video sled during BGR expeditions to the CCZ in the Pacific.

REFERENCES

[1] Hanumant Singh, Ali Can, Ryan Eustice, Steve Lerner, Neil McPhee, and Chris Roman. Seabed auv offers new platform for high-resolution imaging. *Eos, Transactions American Geophysical Union*, 85(31):289–296, 2004.

[2] Olaf Pfannkuche and Peter Linke. Geomar landers as long-term deep-sea observatories. *Sea Technology*, 44(9):50–55, 2003.

[3] T. Soltwedel, E. Bauerfeind, M. Bergmann, N. Budaeva, E. Hoste, N. Jaeckisch, K. v. Juterzenka, J. Matthiessen, V. Mokievsky, E.-M. Nöthig, N. Queric, B. Sablotny, E. Sauter, I. Schewe, B. Urban-Malinga, J. Wegner, M. Wlodarska-Kowalczuk, and M. Klages. HAUSGARTEN: multi-disciplinary investigations at a deep-sea, long-term observatory in the Arctic Ocean. *Oceanography*, 18(3):46–61, 2005.

[4] M Leslie, N Scott, E Guillemot, and V Auger. Video acquisition, archiving, annotation and analysis: NEPTUNE Canada's real-time geo-referenced library of deep sea video. In *OCEANS 2010*, pages 1–9. IEEE, 2010.

[5] Kevin E Kohler and Shaun M Gill. Coral point count with Excel extensions (CPCe): A visual basic program for the determination of coral and substrate coverage using random point count methodology. *Computers & Geosciences*, 32(9):1259–1269, 2006.

[6] Jörg Ontrup, Nils Ehnert, Melanie Bergmann, and Tim W Nattkemper. BIIGLE-Web 2.0 enabled labelling and exploring of images from the Arctic deep-sea observatory HAUSGARTEN. In *OCEANS*, pages 1–7. IEEE, 2009.

[7] T. Schoening, D. Bruen, T. Kuhn, and T. W. Nattkemper. Image-based marine resource exploration and biodiversity assessment with mamas (marine data asset management and analysis system). In *Underwater Mining Institute*, Lisbon, 2014. UMI.

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

[9] Timm Schoening, Melanie Bergmann, Jörg Ontrup, James Taylor, Jennifer Dannheim, Julian Gutt, Autun Purser, and Tim W Nattkemper. Semi-automated image analysis for the assessment of megafaunal densities at the Arctic deep-sea observatory HAUSGARTEN. *PloS one*, 7(6):e38179, 2012.

[10] Kerry L Howell, Ross D Bullimore, and Nicola L Foster. Quality assurance in the identification of deep-sea taxa from video and image analysis: response to Henry and Roberts. *ICES Journal of Marine Science: Journal du Conseil*, pages 899–906, 2014.

[11] Autun Purser, Melanie Bergmann, Tomas Lundälv, Jörg Ontrup, and Tim W Nattkemper. Use of machine-learning algorithms for the automated detection of cold-water coral habitats: a pilot study. *MEPS*, 397:241–251, 2009.

[12] Thomas Kuhn, Carsten Rühlemann, Michael Wiedicke-Hombach, Timm Schoening, and Tim Nattkemper. Application of hydro-acoustic and video data for the exploration of manganese nodule fields. In *ISOPE OMS*, 2013.

[13] Peter A Rona. Resources of the sea floor. *Science*, 299(5607):673–674, 2003.

[14] James R Hein, Tracey A Conrad, and Hubert Staudigel. Seamount mineral deposits: a source of rare metals for high-technology industries. *Oceanography*, 23, 2010.

[15] CL Van Dover, J Aronson, L Pendleton, S Smith, Sophie Arnaud-Haond, D Moreno-Mateos, E Barbier, D Billett, K Bowers, R Danovaro, et al. Ecological restoration in the deep sea: Desiderata. *Marine Policy*, 44:98–106, 2014.

[16] Cindy Lee Van Dover. Mining seafloor massive sulphides and biodiversity: what is at risk? *ICES Journal of Marine Science: Journal du Conseil*, pages 341–348, 2010.

[17] LM Wedding, AM Friedlander, JN Kittinger, L Watling, SD Gaines, M Bennett, SM Hardy, and CR Smith. From principles to practice: a spatial approach to systematic conservation planning in the deep sea. *Proceedings of the Royal Society B: Biological Sciences*, 280(1773):20131684, 2013.

[18] Hartmut Bluhm. Re-establishment of an abyssal megabenthic community after experimental physical disturbance of the seafloor. *Deep Sea Research Part II: Topical Studies in Oceanography*, 48(17-18):3841 – 3868, 2001.

[19] Rahul Sharma, NH Khadge, and S Jai Sankar. Assessing the distribution and abundance of seabed minerals from seafloor photographic data in the central Indian Ocean basin. *International Journal of Remote Sensing*, 34(5):1691–1706, 2013.

[20] Timm Schoening and Tim Wilhelm Nattkemper. Seabed classification using a novel bag-of-prototypes feature representation. Proceedings of the ICPR 2013, 2014.

[21] Norman MacLeod and Phil Culverhouse. Time to automate identification. *Nature*, 467(7312):154–5, 2010.

[22] Autun Purser, Laurenz Thomsen, Chris Barnes, Mairi Best, Ross Chapman, Michael Hofbauer, Maik Menzel, and Hannes Wagner. Temporal and spatial benthic data collection via an internet operated deep sea crawler. *Methods in Oceanography*, 5:1–18, 2013.

[23] Timm Schoening, Thomas Kuhn, Melanie Bergmann, and Tim W Nattkemper. Delphi - fast and adaptive computational laser point detection and visual footprint quantification for arbitrary underwater image collections. *Frontiers in Marine Science*, 2(20), 2015.

[24] Jonas Osterloff, Timm Schoening, Michael Bergmann, Jennifer M Durden, Henry Ruhl, Tim W Nattkemper, et al. Ranking color correction algorithms using cluster indices. In *Computer Vision for Analysis of Underwater Imagery (CVAUI), 2014 ICPR Workshop on*, pages 41–48. IEEE, 2014.

[25] Jörg Ontrup and Helge Ritter. Large-scale data exploration with the hierarchically growing hyperbolic som. *Neural networks*, 19(6):751–761, 2006.