

RESEARCH ARTICLE

Computational Visual Stress Level Analysis of Calcareous Algae Exposed to Sedimentation

Jonas Osterloff^{1*}, Ingunn Nilssen^{2,3}, Ingvar Eide², Marcia Abreu de Oliveira Figueiredo^{4,5}, Frederico Tapajós de Souza Tâmega^{5,6}, Tim W. Nattkemper¹

1 Biodata Mining Group, Faculty of Technology, Bielefeld University, Bielefeld, Germany, **2** Statoil ASA, Research and Technology, N-7005 Trondheim, Norway, **3** Trondhjem Biological Station, Department of Biology, Norwegian University of Science and Technology, N-7491 Trondheim, Norway, **4** Instituto de Pesquisa Jardim Botânico do Rio de Janeiro, CEP 22.460-030, Rio de Janeiro, RJ, Brazil, **5** Instituto Biodiversidade Marinha, 22.793-000, Rio de Janeiro, RJ, Brazil, **6** Instituto de Estudos do Mar Almirante Paulo Moreira, 28.930-000, Arraial do Cabo, RJ, Brazil

* josterlo@cebitec.uni-bielefeld.de



OPEN ACCESS

Citation: Osterloff J, Nilssen I, Eide I, de Oliveira Figueiredo MA, de Souza Tâmega FT, Nattkemper TW (2016) Computational Visual Stress Level Analysis of Calcareous Algae Exposed to Sedimentation. PLoS ONE 11(6): e0157329. doi:10.1371/journal.pone.0157329

Editor: James P. Meador, Northwest Fisheries Science Center, NOAA Fisheries, UNITED STATES

Received: April 15, 2015

Accepted: May 27, 2016

Published: June 10, 2016

Copyright: © 2016 Osterloff et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All images are available under doi:[10.4119/unibi/2775316](https://doi.org/10.4119/unibi/2775316) (<http://doi.org/10.4119/unibi/2775316>).

Funding: Financial support was given by STATOIL Brasil Óleo e Gás Ltda (STATOIL Brazil Oil and Gas LLC) (<http://www.statoil.com/brazil/>). Agência Nacional de Petróleo, Gás Natural e Biocombustíveis - ANP (National Agency of Petroleum, Natural Gas and Biofuels - ANP) (<http://www.anp.gov.br/>), Rio de Janeiro, Brazil, has accepted the Project under the Federal Participation Agreement (FPE). The decision to publish was not influenced by the funders.

Abstract

This paper presents a machine learning based approach for analyses of photos collected from laboratory experiments conducted to assess the potential impact of water-based drill cuttings on deep-water rhodolith-forming calcareous algae. This pilot study uses imaging technology to quantify and monitor the stress levels of the calcareous algae *Mesophyllum engelhartii* (Foslie) Adey caused by various degrees of light exposure, flow intensity and amount of sediment. A machine learning based algorithm was applied to assess the temporal variation of the calcareous algae size (\sim mass) and color automatically. Measured size and color were correlated to the photosynthetic efficiency (maximum quantum yield of charge separation in photosystem II, $\Phi_{PSII_{max}}$) and degree of sediment coverage using multivariate regression. The multivariate regression showed correlations between time and calcareous algae sizes, as well as correlations between fluorescence and calcareous algae colors.

Introduction

Increasing anthropogenic activities in marine areas call for a holistic environmental monitoring approach using new and more sophisticated models and sensor systems that enable predictions and measurement of impact on organisms of interest [1]. One important input to modeling is knowledge generated through laboratory experiments, determining various levels of stress on the organisms of interest. A common way to measure stress is through visual documentation of behavioral changes, change in abundance [2–4], change of the spectral response [5] or change of the visual appearance [6, 7] of organisms.

The investigated calcareous algae species, *Mesophyllum engelhartii* (Foslie) Adey, belongs to the *Phylum Rhodophyta* under the Order *Hapalidiales*. Calcareous algae play an important ecological role in coastal habitats [8–10] contributing to the formation of rhodoliths. Rhodoliths are made from dead calcareous algae and other calcifying organisms [11]. These multi-

Competing Interests: The authors of this manuscript have read the journal's policy and have the following competing interests: Two of the authors, Ingunn Nilssen and Ingvar Eide are employees in Statoil, Research and Technology, Trondheim, Norway. They have contributed scientifically to the manuscript and have been involved in study design and data analyses. The decision to publish was not influenced by the funders; on the contrary, the intention and a prerequisite have been to publish the work in peer reviewed scientific journals and release the data.

spherical structures are creating a habitat for other organisms living on, between and in the structures [12]. Calcareous algae are found down to about 250 meters water depth [13–15] and the largest occurrence of rhodolith beds are found in the southwest Atlantic on the Brazilian continental shelf [16–18]. These algae communities may be disturbed by natural sedimentation and/or sedimentation from anthropogenic activities such as fish-trawling, mining [19] and discharges of drill cuttings from oil and gas drilling activities [20].

In a previous study the impact from sedimentation of drill cuttings on live calcareous algae was studied by measuring sediment coverage (SC) and photosynthetic efficiency (P) [6]. The study was part of the Peregrino Environmental Monitoring Calcareous Algae Project (PEMCA). Rhodoliths, partly covered with calcareous algae, were collected from the Peregrino oil field off the coast of Brazil. The impact of sedimentation was investigated in a laboratory flow-through system, varying light exposure (\mathcal{L}), flow rate (\mathcal{F}), amount of sediment (S) and time (T). Photosynthetic efficiency (P) was measured as maximum quantum yield of charge separation in photosystem II, $\Phi_{\text{PSII,max}}$. These results have been published elsewhere [6].

The present paper describes a new method to extend the observed impact variables (SC , P) by measurements of size and color of calcareous algae in digital photos recorded during the experiments.

Using photos for environmental monitoring generates a large number of images implying that a manual evaluation is extremely demanding regarding time and resources and requires a careful and professional execution, as for instance outlined in [21].

To support the evaluation and labeling of the photos performed by marine biology experts different software tools have been proposed such as Coral Point Count with Excel extensions (CPCe) [22]. Recently, web-based image annotation and labeling systems have been proposed to support web-based sharing of photos and collaboration, such as BIIGLE (Benthic Image Indexing, Graphical Labeling and Exploration) [23], which was successfully applied in [4, 24–28], and CoralNet [29], the latter one providing for instance an automated point classification for corals as well.

Since humans have a limited capability to quantify visual features, such as color change, in an objective way, we propose a machine learning based approach to compute size and color of the calcareous algae from photos recorded at different time points. A machine learning algorithm is trained with a number of manual image annotations. The trained classifier is applied for the full automatic segmentation of the calcareous algae, enabling the quantification of the calcareous algae size and color over time. To learn the classification function a H^2SOM (Hierarchical Hyperbolic Self-Organizing Map) [30] algorithm was applied. This unsupervised learning algorithm has previously been used for cold-water coral segmentation in ROV video frames [24] and for poly-metallic nodule segmentation [31, 32]. Nevertheless, the integration of the basic algorithm into an image analysis pipeline needed to be modified substantially regarding pre-processing, feature computation, pixel classification function and post-processing.

A simplified overview of the whole impact quantification process is given in Fig 1.

Material

Samples of live calcareous algae *Mesophyllum engelhartii* (Foslie) Adey were collected from 94–103m water depth at the Peregrino oil production field (23°13'28.34"S, 41°50'55.73"W) located off the Brazilian Atlantic coast [11]. Light intensity (\mathcal{L}), flow rate (\mathcal{F}) and sediment amount (S) could be varied independently as described in [6]. In order to optimize the experiment, the three predictor variables (\mathcal{L} , \mathcal{F} , S) were combined according to a Central Composite Face (CCF) design with center point [33]. This enables regression modeling to describe linear

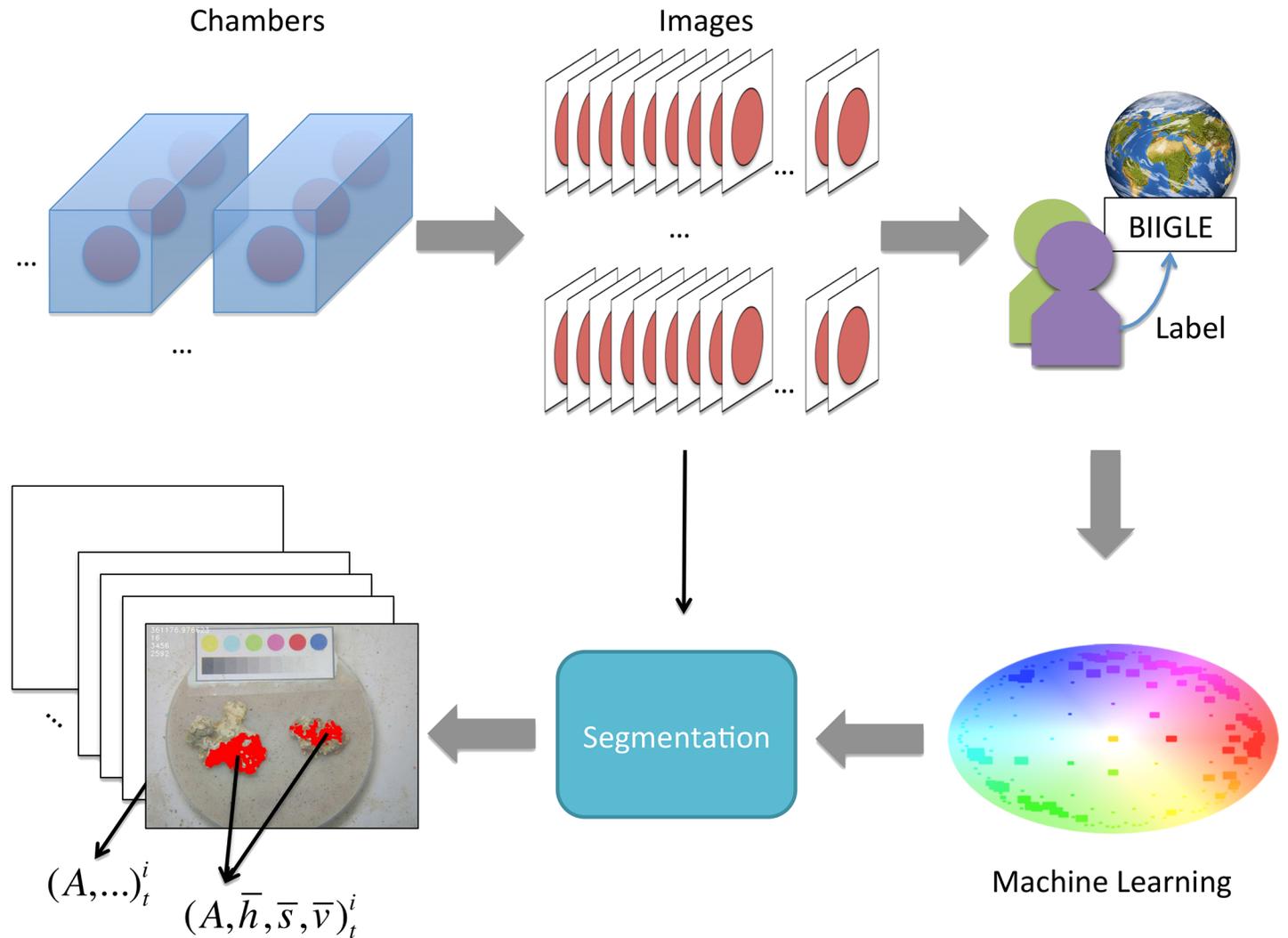


Fig 1. A simplified illustration of the impact quantification process. The process is starting on the upper left with the laboratory experiment, continuing with the image recording, the labeling by the experts, the application of the unsupervised machine learning algorithm (see Machine Learning section for details) and ending with the image segmentation and the measuring of size and color of the segmented area.

doi:10.1371/journal.pone.0157329.g001

and non-linear relationships and interactions between the variables. The full factorial design gives 15 combinations of \mathcal{L} , \mathcal{F} , and \mathcal{S} . Two of these combinations were duplicated and two were triplicated (i.e. technical replicates). Using the 15 combinations and technical replicates resulted in a total of 21 experiments which were carried out in 21 chambers (Table 1). In addition controls were kept in complete darkness and without sediment. The area of the chamber bottom was approximately 0.06m^2 and the height of was 0.1 m resulting in a volume of approximately 6 L. The light levels (\mathcal{L}) used in the exposure studies were 3, 6.6, or $10 \mu\text{mol m}^{-2} \text{s}^{-1}$, the latter value corresponding to measured light conditions at the sea floor on the Peregrino field. The flow rates (\mathcal{F}) used were 0.04, 0.07, or 0.09 m s^{-1} . The amounts of sediment (\mathcal{S}), applied at the beginning of the experiment (T_0), were 600, 900, or 1200 g per chamber. The chosen amounts resulted in a sediment coverage (SC) of the calcareous algae ranging from uncovered to completely covered, implying that also non- and very low sediment covered calcareous algae are included in the analysis. Six parallels (i.e. biological replicates) of rhodoliths

Table 1. Chamber configurations for the CCF design.

Chamber	\mathcal{L} [$\frac{\mu\text{mol}}{\text{m}^2\text{s}}$]	\mathcal{F} [$\frac{\text{m}}{\text{s}}$]	S [g]
1	3.0	0.04	600
2	10.0	0.09	600
3	3.0	0.09	600
4	10.0	0.04	600
5	3.0	0.04	1200
6	10.0	0.04	1200
7	3.0	0.09	1200
8	10.0	0.09	1200
9	3.0	0.07	900
10	3.0	0.07	900
11	10.0	0.07	900
12	6.6	0.04	900
13	6.6	0.04	900
14	6.6	0.04	900
15	6.6	0.09	900
16	6.6	0.09	900
17	6.6	0.09	900
18	6.6	0.07	600
19	6.6	0.07	1200
20	6.6	0.07	900
21	6.6	0.07	900

The configurations for the 21 chambers for light intensity (\mathcal{L}), flow rate (\mathcal{F}), and sediment amount (S) are displayed. Replicated configurations are grouped.

doi:10.1371/journal.pone.0157329.t001

covered with healthy calcareous algae were placed in each chamber, so each of the 21 experiments were conducted with six samples.

In addition to the measurements reported in [6], photos of the calcareous algae samples were taken for purpose of the present paper during the weekly examination. For image-acquisition, the calcareous algae rodoliths were removed from their chambers and manually placed in a transparent, water-filled cylinder together with a color reference plate (Figs 2 and 3A). The reference plate, essential for a comparative study of the color over time, can be seen in the upper part of the images (Fig 3 upper left). The photos were taken with a Canon PowerShot G10, which was placed on a tripod to control the distance (15 cm) between camera and samples. The camera was in automatic mode with macro function and internal flash on. To eliminate light refractions and reflections samples and camera, the latter protected by a waterproof housing, were placed underwater inside the cylinder (Fig 2). The digital images had a resolution of 3456×2592 pixels. Each week 63 images (21 chambers \times 3 images, 2 samples per photo), were taken summing up to $N = 630$ images for the whole experiment (T0, . . . , T9).

In order to collect a set of image annotations for training and testing the machine learning algorithm, the recorded images were uploaded to BIIGLE. With this online image annotation tool, biological experts from the Instituto Biodiversidade Marinha were manually labeling photo regions (pixel) as “live” calcareous algae (2404 labels), “stressed” calcareous algae (2834 labels), “dead” calcareous algae (4358 labels) and “bare substratum” (43 labels) on all $N = 630$ images. For the annotation, the experts were allowed to select single point label or customizable frame labels (Fig 4). These labels provided examples for the later machine learning training

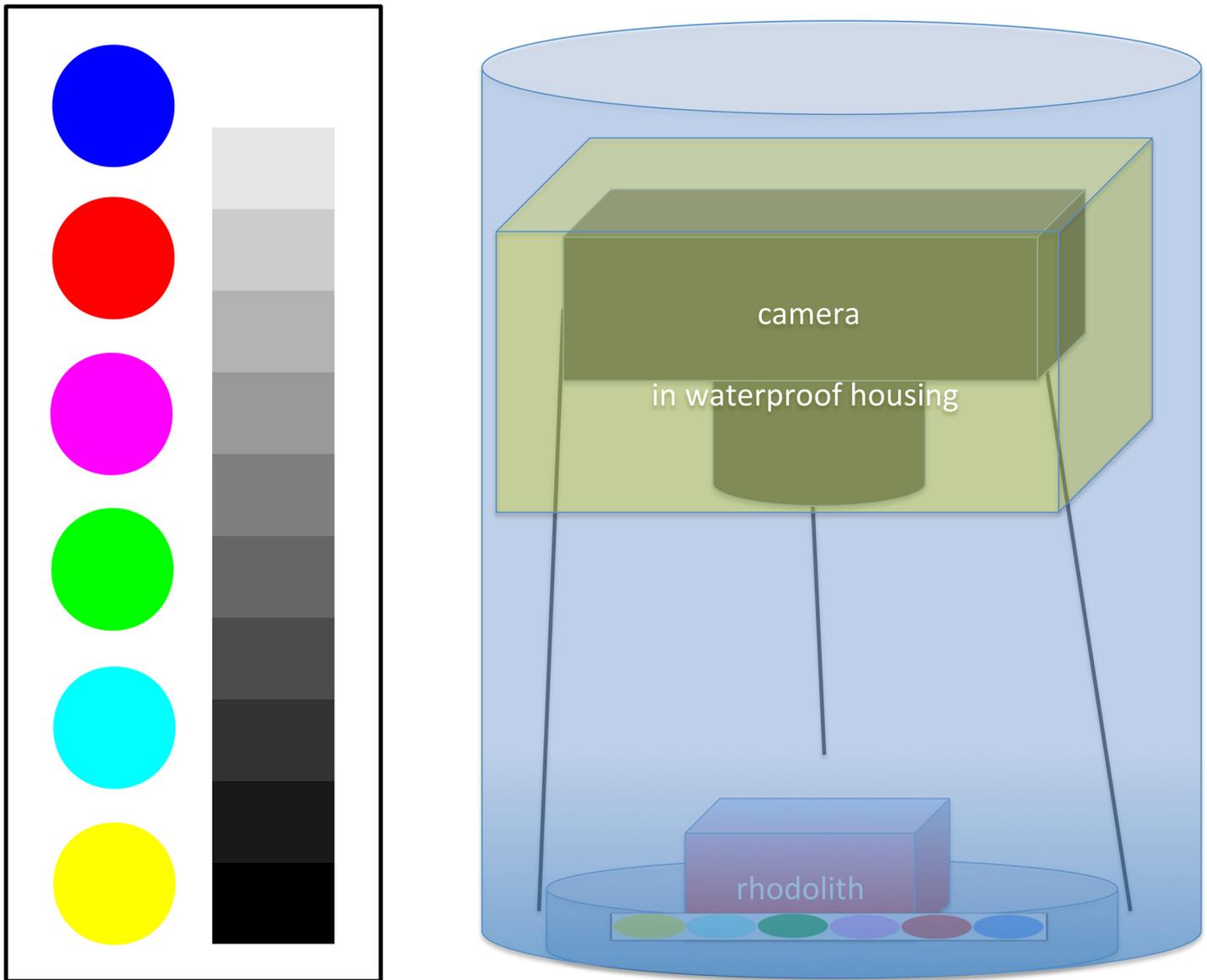


Fig 2. Illustration of color reference plate and camera setup. On the left the original color reference plate, used to obtain color constancy within the experiment, is displayed. Colors are referring to the RGB triplets (top-down): (0,0,255), (255,0,0), (255,0,255), (0,255,0), (0,255,255), (255,255,0). On the right a schematic layout of the underwater camera setup is presented, consisting of a water filled cylinder containing the sample, the color reference plate and the camera in the waterproof housing fixed on a tripod.

doi:10.1371/journal.pone.0157329.g002

step (see Machine Learning section below). The labeling strategy was chosen to reduce the time needed for the labeling and simultaneously cover the whole variation between and within each label category over time. The “live” and “stressed” calcareous algae labels represent the so called “positive” class in the segmentation and the “dead”- and “bare substratum”-labels represent the so called “negative” class. All images and labels can be accessed via BIIGLE (<https://ani.cebitec.uni-bielefeld.de/biigle/>). To browse the data, a login with username **pemca** and password **pemca** is required. The data is stored in the project **Pemca3rd**, with the transects T_0, \dots, T_9 . Image material is also published under DOI [10.4119/unibi/2775316](https://doi.org/10.4119/unibi/2775316).

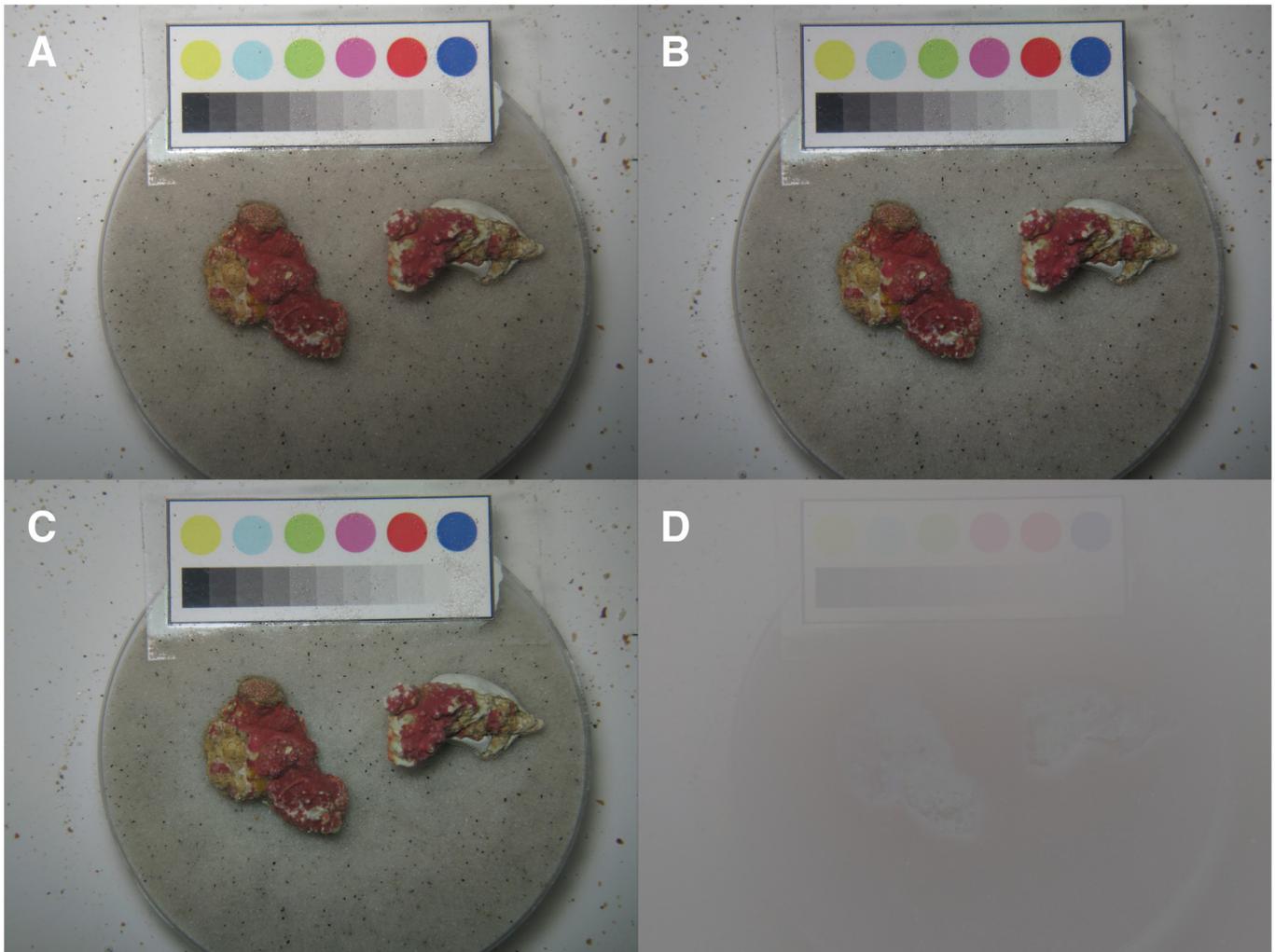


Fig 3. Example of an image recorded during the experiment completion and its different processing results of the pre-processing. The original image (A), the result of the illumination correction (B), the result of the color correction (C) and the difference image (D) are displayed. In the original image (A) the color reference plate is located in the upper part, providing input for the pre-processing. The two calcareous algae samples are located below the reference plate. Comparing the result image from the illumination correction (see Illumination Correction section for details) with the original image a slight decrease in brightness can be noticed. Due to the stable illumination conditions during image recording only little differences between the images can be recognized visually. The color correction (see Reference Plate Detection, Color and Zoom Correction section for details) is changing the colors very little as only minor color shifts appear. Most of the visible differences in the difference image are caused by the illumination correction.

doi:10.1371/journal.pone.0157329.g003

Methods

The permission to collect the calcareous algae samples for the experimental study described in the Material section was granted by the Chico Mendes Institute for Biodiversity Conservation SISBIO license n 20820-2 and 20826-1. The experimental study did not involve endangered or protected species.

The basic idea of the computational calcareous algae segmentation is to classify each pixel in each image to be i) “live” or “stressed” calcareous algae or ii) other, including “dead” calcareous algae. A pixel classifier is trained using data collected from the annotations provided by the experts (see [Material](#) section). From the pixel-wise classification, the regions of “live”/“stressed” calcareous algae are determined for each sample at each time point. The results are analyzed in

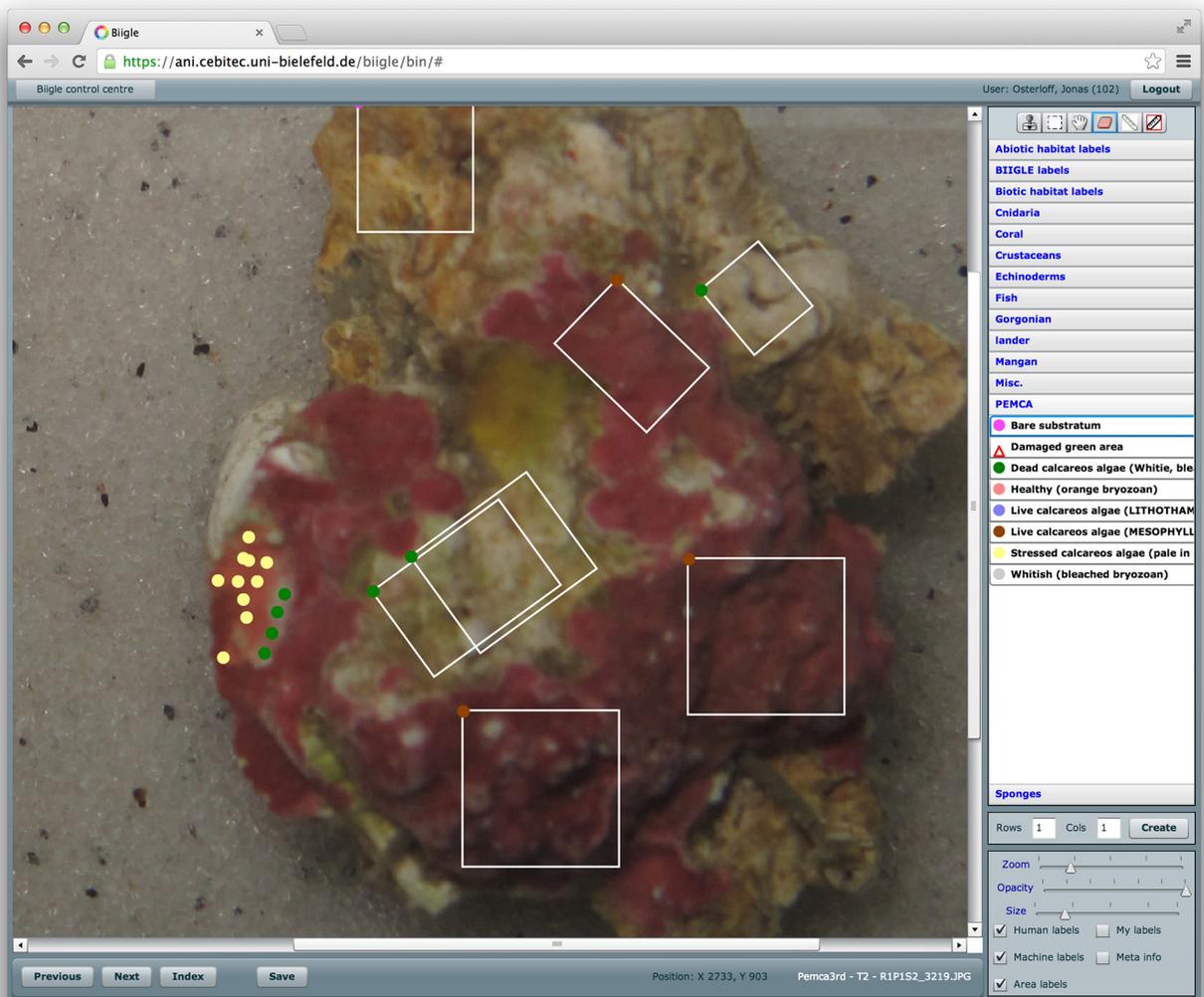


Fig 4. A screenshot of the BIIGLE system in a web-browser. The images of the calcareous algae were examined and labeled using the BIIGLE system. Images can also be zoomed to examine more details. The experts were allowed to select single point label or customizable frame labels. The single point labels are represented as filled colored circles, the customizable frame labels are represented as white outlined rectangles with a filled colored circle in the upper left corner of the individual rectangle. The colors of the filled circles indicate the class of the individual label. Red is representing “live” calcareous algae, yellow “stressed” calcareous algae, green “dead” calcareous algae and Pink “bare substratum”.

doi:10.1371/journal.pone.0157329.g004

combination with the previously measured parameters photosynthetic efficiency (P) and sediment coverage (SC) using multivariate analysis. The individual steps for image processing, pixel classification and final data analysis are described in detail in the following sections.

Pre-Processing

Before the feature extraction and the machine learning can be applied, the recorded images need to be pre-processed.

Illumination Correction. In order to compensate illumination fluctuations in the images the *local space average color scaling* [34] is used on each image individually. The local space average color image is computed using a Gaussian blurring with $\sigma = 0.093 \cdot \max(I_{\text{height}}, I_{\text{width}})$.

As the images were captured in sRGB [35], which assumes a gamma correction of 1/2.2, the images are gamma corrected and then divided pixel- and channel-wise (c) by the corresponding Gaussian filtered image $G^{(n)}$:

$$\hat{I}_{x,y,c}^{(n)} = \left(\frac{(I_{x,y,c}^{(n)})^{2.2}}{\tau \cdot G_{x,y,c}^{(n)}} \right)^{1/2.2} \tag{1}$$

τ is a scaling factor and set to $\tau = 2$ in the whole study. An example for an illumination corrected image is displayed in Fig 3 (upper right).

Reference Plate Detection, Color and Zoom Correction. The first step in the color correction is the detection of the color circles in the reference plate using a Hough circle transformation [36]. The Hough transform computes a list of circle candidates and those that show a low color variance inside are selected taking the spatial layout of the color plate into account as well. The identified six color circles C_l ($l = 1, \dots, 6$) of each $\hat{I}^{(n)}$ are then used to calculate one RGB color triplet $\mathbf{c}_l^{(n)}$ for each circle over all its pixels $p_{(x,y),c}^{(n)}$.

$$\mathbf{c}_l^{(n)} = \begin{pmatrix} \text{median}_{(x,y) \in \text{circle}_l} (p_{(x,y),\text{red}}^{(n)}) \\ \text{median}_{(x,y) \in \text{circle}_l} (p_{(x,y),\text{green}}^{(n)}) \\ \text{median}_{(x,y) \in \text{circle}_l} (p_{(x,y),\text{blue}}^{(n)}) \end{pmatrix} \tag{2}$$

In addition, overall median color references are computed for each of the six color circles (Eq 3) of all images (Fig 5).

$$\bar{\mathbf{c}}_l = \begin{pmatrix} \text{median}_n (\mathbf{c}_{l,\text{red}}^{(n)}) \\ \text{median}_n (\mathbf{c}_{l,\text{green}}^{(n)}) \\ \text{median}_n (\mathbf{c}_{l,\text{blue}}^{(n)}) \end{pmatrix} \tag{3}$$

For each image $\hat{I}^{(n)}$ and color channel $c \in \{\text{red, green, blue}\}$, the individual, channel-wise gamma correction is calculated using the color RGB triplet $\mathbf{c}_l^{(n)}$ (Eq 2) of the individual image

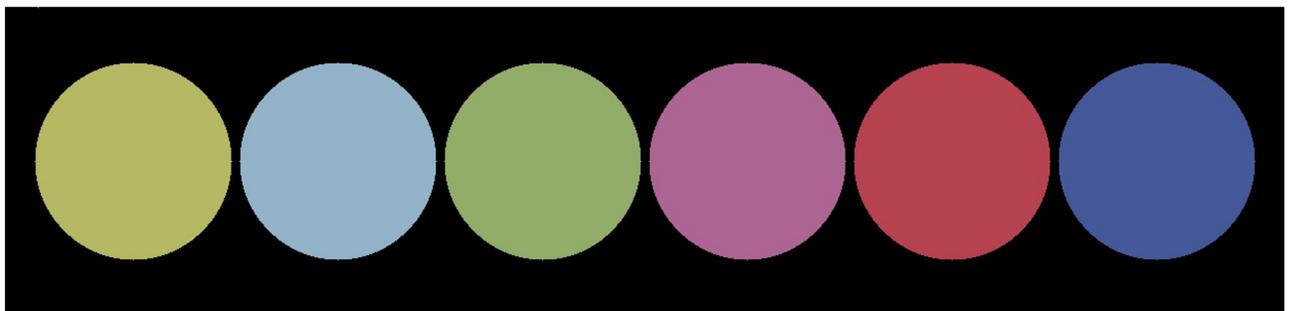


Fig 5. The average color reference plate. This reference is computed by averaging over the detected reference plates from all images used in this study (see text for details). Colors are referring the RGB triplets (left to right) (181, 185, 100), (147, 179, 200), (146, 173, 106), (173, 101, 147), (183, 67, 80), and (69, 88, 154).

doi:10.1371/journal.pone.0157329.g005

and the overall median color references (Eq 3).

$$\gamma_c^{(n)} = \text{median}_I \left(\frac{\log \left(\frac{c_{I,\text{red}}^{(n)}}{255} \right)}{\log \left(\frac{c_{I,c}^{(n)}}{255} \right)} \right) \tag{4}$$

This image specific gamma correction is applied channel-wise to all the images resulting in a new set of color corrected images (Fig 3 lower left)

$$J_{(x,y),c}^{(n)} = \left(\hat{I}_{(x,y),c}^{(n)} \right)^{\gamma_c^{(n)}} \tag{5}$$

A difference image $(I_{(x,y,c)}^{(n)} - J_{(x,y,c)}^{(n)}) / 2 + 128$ is presented in Fig 3 (lower right) to demonstrate the impact of the described illumination and color correction.

The layout of the reference plate is also used to correct zoom variations, which occurred for a significant number of images. The average distance \bar{d} between the neighboring color circles (e.g. green and magenta) on each plate is evaluated from the five Euclidian circle distances $d_{i, i+1}$

$$\bar{d} = \frac{1}{N} \sum_n \frac{1}{5} \sum_{i=1}^5 d_{(i,i+1)}^{(n)} \tag{6}$$

The difference between the average distance and the observed individual distances in image $\{J^{(n)}\}$ is used to rescale each image individually.

Segmentation

Unsupervised learning is applied to learn a vector quantization of the image feature space (see Machine Learning section). Vector quantization fits a set of so called prototype vectors iteratively to a data set of feature vectors from the images describing pixel features. After the training step is finished, each prototype represents a group of pixels sharing similar features, i.e. a dense cloud in the feature space. The prototypes are assigned to the classes “live / stressed calcareous algae” and “other”, including “dead calcareous algae” and “bare substratum”, and are applied to classify all image pixel (see Prototype Label Mapping section). For the final segmentation result the classified image is post-processed (see Post Processing section).

Feature Extraction. One important step in vector quantization-based image segmentation is the selection of appropriate features and the collection of a training set. To collect a training set for the vector quantization learning, a selection of image pixels $\{p_i\}$ from a subset of all N images are chosen and their feature vectors $\{\mathbf{x}^{(i)}\}$ are computed. Here, Median RGB values extracted on a 11×11 pixel neighborhood $(\eta_{11}^{(i)})$ for each color channel, are used as features since we observed such a representation to be sufficient for the initial task of calcareous algae segmentation by pixel classification.

$$\mathbf{x}^{(i)} = \begin{pmatrix} \text{median}_{(x,y) \in \eta_{11}^{(i)}} \left(p_{(x,y),\text{red}} \right) \\ \text{median}_{(x,y) \in \eta_{11}^{(i)}} \left(p_{(x,y),\text{green}} \right) \\ \text{median}_{(x,y) \in \eta_{11}^{(i)}} \left(p_{(x,y),\text{blue}} \right) \end{pmatrix} \tag{7}$$

To reduce the computation time, features are only extracted for pixels in a 2×2 grid resolution collecting a training set of about 2.5 million feature vectors $\{\mathbf{x}^{(i)}\}$.

Machine Learning. To perform vector quantization on the training data the unsupervised machine learning algorithm H²SOM [30] is applied.

H²SOMs with different parameter settings, like the number of prototypes, were trained and the best parameter set was determined using a customized criterion described in (S1 Text). A training set of feature vectors generated from 12.5% of the pixels chosen from a small subset (2%) of the entire image collection is used to construct the H²SOM with the best segmentation results. The H²SOM is trained using $30 \times |\{\mathbf{x}^{(i)}\}|$ iterations, an exponentially decreasing learning rate of $\alpha = [0.99, 0.1]$, an eight-neighbor topology and three rings and therefore is consisting of 161 prototypes.

Prototype Label Mapping. The next step is to build a pixel classifier out of the trained H²SOM, using the feature vectors $\mathbf{x}^{(p)}$ of all images $f^{(n)}$. We distinguish between feature vectors $\mathbf{x}^{(i)}$ used for the H²SOM training and feature vectors $\mathbf{x}^{(p)}$ from all images, but of course both are computed as described in Eq 7. The expert labels “live” and “stressed” from BIIGLE are used to build the classifier as follows. In general, we have observed that a differentiation between the categories “live” and “stressed” is problematic regarding the reproducibility of the results of the visual assessment. As a consequence a discrimination between these two categories has been neglected. The fused category is referred to as “live calcareous algae”.

First, a label y_p is assigned to each feature vector $\mathbf{x}^{(p)}$

$$y_p = \begin{cases} 1, & \text{if the pixel was labeled as “live”/“stressed” calcareous algae} \\ 0, & \text{else} \end{cases} \quad (8)$$

Second, each feature vector $\mathbf{x}^{(p)}$ is assigned to one of the H²SOM prototypes $\{\mathbf{u}^{(j)}\}$ ($j = 1..161$) that represents the features of $\mathbf{x}^{(p)}$ in the best possible way. This prototype is selected by the so called Best Matching Unit (BMU) criterion, that selects the prototype with the minimal Euclidean distance to $\mathbf{x}^{(p)}$ in the feature space. Next, all cluster prototypes are investigated in order to classify them. The prototypes constitute a Voronoi tessellation of the feature space

$$V_j = \{\mathbf{x}^{(p)} | j = \underset{j'}{\operatorname{argmin}}\{\operatorname{euclid}(\mathbf{u}^{(j')}, \mathbf{x}^{(p)})\}\} \quad (9)$$

with $q_j = |V_j|$ as the number of feature vectors that have been assigned to $\mathbf{u}^{(j)}$ and therefore are located in V_j . In each Voronoi cell V_j we now analyze how many feature vectors stem from pixel, labelled as “live” or “stressed”:

$$V_j^+ = \{\mathbf{x}^{(p)} | j = \underset{j'}{\operatorname{argmin}}\{\operatorname{euclid}(\mathbf{u}^{(j')}, \mathbf{x}^{(p)})\} \wedge y_p = 1\}, \quad (10)$$

with $q_j^+ = |V_j^+|$ as the number of feature vectors in V_j that are labeled as “live”/“stressed” calcareous algae. The relation $s_j = q_j^+ / q_j$ quantifies how well a prototype $\mathbf{u}^{(j)}$ resembles the features of live calcareous algae. Since we do not expect all $\mathbf{x}^{(p)}$ with $y_p = 1$ to be perfect representatives, only prototypes $\mathbf{u}^{(j)}$ with a high s_j will represent the positive class (presence of live calcareous algae).

Next the relation values $\{s_j\}$ are sorted in descending order:

$$s_j \rightarrow s_i, \text{ with } s_i > s_{i-1} \forall i. \quad (11)$$

From this list, the first L prototypes $u^{(l)}$ are chosen as a model for typical live calcareous algae color features. L is chosen so at least 80% of the entire positive training data (i.e. $y_i = 1$) is covered. This set of the first 16 “most live calcareous algae-like” prototypes is used for the pixel

classification based segmentation in all images. In each image, each pixel p is mapped to each feature vector $\mathbf{x}^{(p)}$. The BMU of the feature vector is determined and if this unit is a live calcareous algae-like prototype, as described above, the pixel is labeled as live calcareous algae ($L(\mathbf{x}^{(p)}) = 1$) or otherwise as background ($L(\mathbf{x}^{(p)}) = 0$).

Post Processing. By classifying all feature vectors $\mathbf{x}^{(p)}$, binary images are created showing white pixels on live calcareous algae classified positions and black pixels otherwise. These binary images are post-processed with an opening filter mask [37] of a size 10×10 pixels to remove small areas of false positive classifications. The post-processed binary images represent the final segmentation result.

Multivariate Data Analysis

To analyze the possible biological impact of sedimentation on calcareous algae, the size and the color of the segmented areas in the images are computed. (Fig 6). The size (A) of the live calcareous algae area is measured as the amount of segmented pixels per image $J^{(n)}$. It is weighted by the ratio of the average color circle distance of circle centers \bar{d} and the average circle center distance \bar{d}_n of the color reference plate from the Image $J^{(n)}$.

$$A = \text{size} \times \frac{\bar{d}}{\bar{d}_n} \tag{12}$$

The relative size $\hat{A} = \frac{A}{A_0}$, with A_0 being the live calcareous algae size at the first measurement, is finally used as one variable. The HSV (hue, saturation, value) color space is used for measuring the color. Note that we use a different color space here than the one we used for the segmentation. Averages of hue (averaging is applied with respect that it is a circular quantity) (\bar{h}), saturation (\bar{s}) and value (\bar{v}) of all segmented pixels per image $J^{(n)}$ are computed of the segmented pixels. These four variables are combined to one measurement for the image n :

$$\mathbf{f}^{(n)} = (\hat{A}, \bar{h}, \bar{s}, \bar{v})^T \tag{13}$$

All calcareous algae samples from the same chamber are exposed to the same treatment parameters. On each image two samples were captured. To avoid the need of a colocation of the samples as well as taking into account that all rhodoliths are different, averages of size and color within each chamber are used. The dataset can be found in [S1 Table](#).

Multivariate regression is performed with partial least squares (PLS) [38, 39] on the live calcareous algae area measurements $\mathbf{f}^{(n)}$, using Unscrambler X 10.3 (CAMO Software, Oslo, Norway) to correlate the predictor variables (X matrix) to the measured responses (Y matrix). The previously published data (first experimental series with *M. engelhartii* in [6]) on P and SC are included in the data matrix for the regression analysis. To match $\mathbf{f}^{(n)}$ also data for P and SC were averaged chamber wise (six parallels). The final data matrix has 168 rows (21 chambers \times 8 time-points) (observations), four columns for the four predictor variables (\mathcal{L} , \mathcal{F} , \mathcal{S} and \mathcal{T}), and six columns for the response variables (\hat{A} , \bar{h} , \bar{s} , \bar{v} , P , SC). Prior to the multivariate data analysis, the data are mean centered and scaled to unit variance. The models were evaluated with respect to explained variance, and goodness of fit and prediction, the latter obtained after cross validation [40, 41].

In addition to the PLS regression, a traditional statistical analysis was conducted by performing ANOVA, MANOVA and pairwise correlation analysis. These results are presented in the supplementary (Table A, Table B, Table C in [S2 Text](#) and Fig A, B, C in [S3 Text](#)). In the following we will concentrate on the results of the PLS regression as it takes all variables into

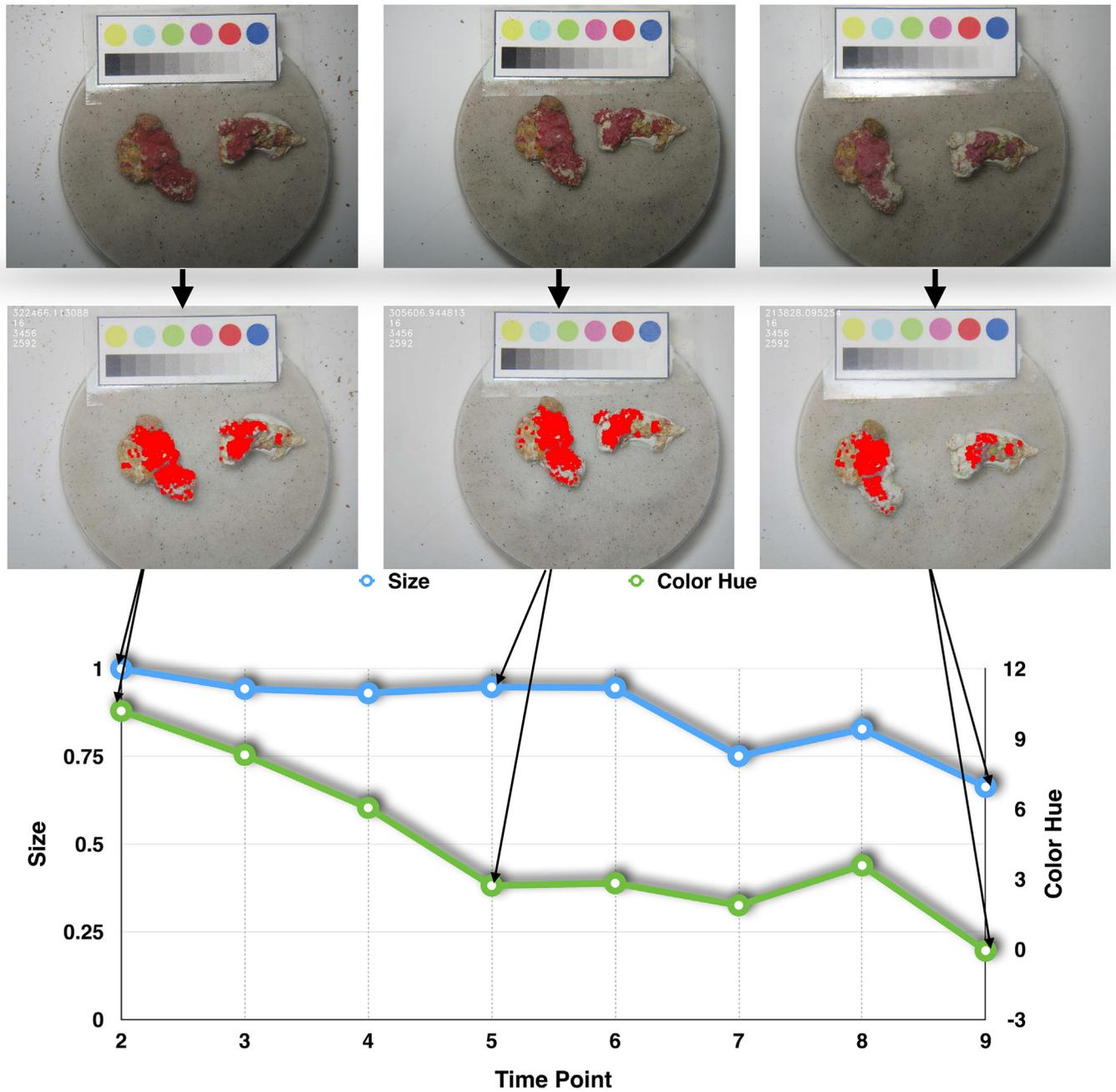


Fig 6. Example of one pair of calcareous algae samples. In the top row the original images recorded at T2, T5 and T9 are shown. In the middle row the segmentation results are highlighted and in the lower row the live calcareous algae relative size (\hat{A}) and hue (\hat{h}), starting from 10 (slightly orange red) to 0 (red), are presented. The samples were exposed to $S = 900 \text{ g}$ (a partially covered sample), $\mathcal{L} = 6.6 \mu\text{mol m}^{-2} \text{ s}^{-1}$ and $\mathcal{F} = 0.07 \text{ m s}^{-1}$.

doi:10.1371/journal.pone.0157329.g006

consideration simultaneously and gives a better understanding of the relative importance and the significance of all variables.

Results

Pre-Processing

Fluctuations in sample illumination are a well-known problem in computational image analysis and the majority of the $N = 630$ images $I^{(n)}$ ($n = 1, \dots, N$) suffered at least partially from it. The procedures described above compensated this effect (Fig 3(B)).

In 15% of the images used in the analysis, the reference color circles could not be detected automatically. In these images the circles were marked manually or only a subset of detected circles was used for the color correction. Although reflections were avoided on the calcareous algae samples, serious reflection problems were observed on the color reference plates in some of the images. The images from T0 and T1 had to be excluded from further analysis because of this problem.

Visual inspection showed that the channel-wise gamma correction produced robust results. The pre-processing stages described above could be computed all together in about three minutes per image.

Segmentation

The segmentation results were reviewed by comparing the human expert labels with the segmented areas of all images using standard measures for image analysis accuracy assessment like recall (or sensitivity), precision (or positive predictive value) and false positive rate (FPR). The recall is the percentage of all pixels labeled “live” or “stressed” $p^{(+)}$ that have been segmented correctly as live calcareous algae by the algorithm.

$$\text{recall} = \frac{|\{p^{(+)}\} \cap \{p^{(s)}\}|}{|\{p^{(+)}\}|} \quad (14)$$

The precision is the percentage of all pixels segmented as live calcareous algae ($p^{(s)}$) that have been segmented correctly by the algorithm.

$$\text{precision} = \frac{|\{p^{(+)}\} \cap \{p^{(s)}\}|}{|\{p^{(s)}\}|} \quad (15)$$

From the segmentation results we computed a recall value of 0.89 and a precision of 0.03.

Since the manually labeling did not attempt to measure the precise live calcareous algae extent in each image, also the “dead” and “bare substratum” labels were used to review the segmentation results. The false positive rate (FPR) was used to measure the number of pixels that were segmented to be “live” or “stressed” $p^{(s)}$ but were labeled “dead” or “bare substratum” $p^{(-)}$.

$$\text{FPR} = \frac{|\{p^{(-)}\} \cap \{p^{(s)}\}|}{|\{p^{(-)}\}|} \quad (16)$$

The FPR was very low compared to the recall. Looking at the false positive rates in detail it should be noted that the FPR for the “bare substratum” is higher than the one of the “dead” labels (Table 2). Segmentation results are also displayed in Fig 7.

Multivariate Data Analysis

The best PLS regression model was obtained after expansion with one square term (S^2) in the polynomial describing the relationship between predictors (\mathcal{L} , \mathcal{F} , \mathcal{S} and \mathcal{T}) and responses

Table 2. Results for the evaluation of the segmentation.

Recall	Prec.	FPR		
		“dead”∪“bare”	“dead”	“bare”
0.89	0.03	0.02	0.02	0.05

The very low values for precision were expected, since a realistic quantification of the FP is not possible in our dataset (see text for details). The low FPR though indicates that the segmentation generates low FP in the “dead” or “bare substratum” labeled regions.

doi:10.1371/journal.pone.0157329.t002

(\hat{A} , \bar{h} , \bar{s} , \bar{v} , P , SC). One significant outlier was identified and removed from the data set, as it appeared that the segmentation failed in the corresponding image (no “live” or “stressed” calcareous algae were detected). Fig 8 shows the correlation loading plot obtained from the PLS-regression. The outer circle represents 100% explained variance and the inner circle 50% explained variance for the different variables.

SC and P , were both positively and negatively (respectively) correlated with amount of sediment (\mathcal{S}) with excellent goodness of fit and prediction. Also the HSV-variables (\bar{h} , \bar{s} , \bar{v}) extracted with our approach, were correlated with \mathcal{S} . Goodness of fit and prediction, the latter obtained after cross validation, are summarized in Table 3. Furthermore, there were inter-correlations between the response variables: the HSV-variables were inter-correlated with SC , and inversely correlated with P . The predictor parameters \mathcal{F} and \mathcal{L} had low influence on the measured responses. Time (T) showed no correlation with P , SC , or HSV, but was well explained and inversely correlated with relative size (\hat{A}). This implies that there are two major directions in the data reflected in the loading plot. The first (horizontal) PLS-component describes \mathcal{S} and measured P , SC and the HSV-variables. The second PLS-component (vertical) describes T and \hat{A} , which are not correlated with the other variables shown in the loading plot.

The apparent inverse correlation between photosynthetic efficiency (P) and the HSV-variables suggests an alternative PLS-regression with HSV-variables as predictors and P as response. Fig 9 shows the resulting loading plot obtained after PLS-regression with linear relations. Goodness of fit and prediction were both 0.82, enabling prediction of photosynthetic efficiency from HSV-values.

Discussion

The use of BIIGLE enabled a flexible and swift exchange of data between the researchers involved in this interdisciplinary study, independent of their geographical location.

The selected strategy for the manual labeling in this study was to apply an “example based” labeling approach, rather than to fully outline the different calcareous algae regions. The rationale behind the chosen strategy was that color changes were expected to be time and/or treatment related and that the use of a “example based” annotation approach (rectangles and points) would balance the time spent on labeling and the amount of training data needed. Due to this approach, no representative ground truth for the negative class, i.e. pixel regions that show no calcareous algae, existed. In principle the negative class could have been constructed by a random selection from the unlabeled regions of the images. However this random selection might have introduced some false-negatives to the negative class. This motivated us to apply an unsupervised learning approach. The prototype label mapping in the H²SOM followed a conservative strategy to classify only 80% of the positive training data as “live”/“stressed” calcareous algae. Despite of this strategy, our segmentation result showed a high

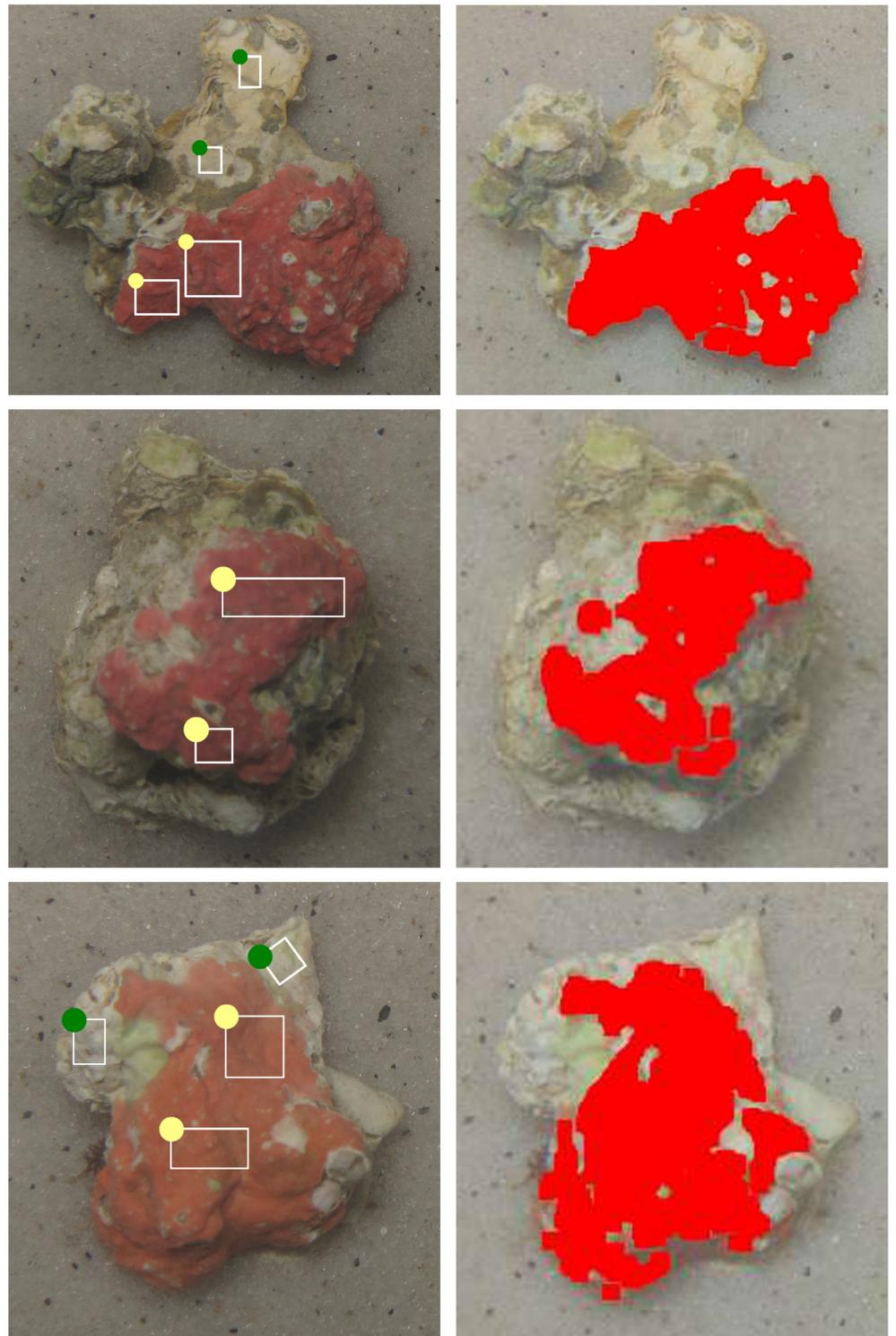


Fig 7. Examples of manually labeled rhodoliths (left) and the corresponding segmented rhodoliths (right). White rectangles with a green dot mark “dead” calcareous algae labeled regions and rectangles with a yellow dot mark “stressed” calcareous algae labeled regions. The pixel regions of “live” or “stressed” calcareous algae are marked red.

doi:10.1371/journal.pone.0157329.g007

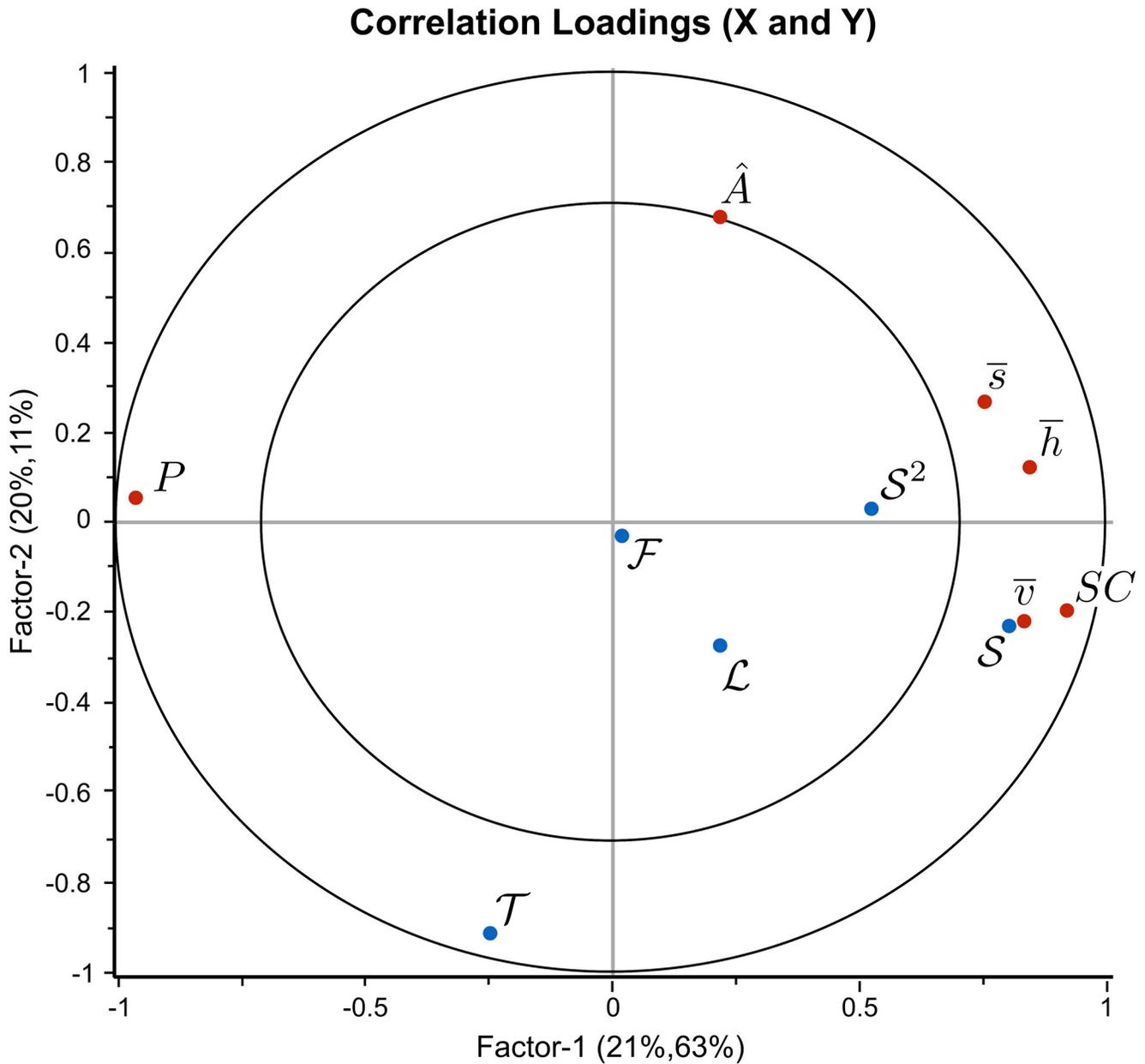


Fig 8. The correlation loading plot obtained after PLS-regression. It was computed with light (\mathcal{L}), flow rate (\mathcal{F}), amount of sediment (S , S^2), and time (\mathcal{T}) as predictors (blue dots), and HSV-variables (\bar{h} , \bar{s} , \bar{v}), relative size (\hat{A}), photosynthetic efficiency (P) and sediment coverage (SC) as responses (red dots). The outer circle represents 100% explained variance and the inner circle 50% explained variance for the different variables.

doi:10.1371/journal.pone.0157329.g008

recall of 0.89. Based on the selected “example based” manual labeling strategy the precision was low, as we were expecting. The rationale behind the chosen strategy was that color changes were expected to be time and/or treatment related and that the use of a “example based” annotation approach (rectangles and points) would balance the time spent on labeling and the amount of training data needed. A posterior visual inspection of the results showed that the selected approach produced many false positives that actually turned out to be true positives. Computing the precision as defined in Eq 15 was therefore not a reasonable assessment for the

Table 3. Goodness of fit and prediction of the PLS-regression.

R^2	\bar{h}	\bar{s}	\bar{v}	\hat{A}	P	SC
fit	0.73	0.64	0.76	0.52	0.93	0.93
pred.	0.72	0.62	0.74	0.50	0.93	0.93

Goodness of fit and prediction for the response variables as function of the four predictor variables \mathcal{L} , \mathcal{F} , \mathcal{S} and \mathcal{T} .

doi:10.1371/journal.pone.0157329.t003

accuracy of our segmentation system (Fig 7). The FPR scores should be low, as an indication of the effect, that almost none of the pixels labeled as “dead” or “bare substratum” were segmented into “live” or “stressed” calcareous algae regions. The FPR for the “bare substratum” is noticeable, but was interpreted as still low and does not seem to have a significant effect on the evaluated colors.

The PLS regression showed that the photosynthetic efficiency is correlated ($R^2 = 0.82$) to the HSV-values, implying that it is possible to predict the photosynthetic efficiency from the color (Fig 9). The segmented images indicate that the area of live calcareous algae decreases with time during the exposure study. However, the correlation of P , SC and the color (\bar{h} , \bar{s} , \bar{v}) with time was low, as goodness of fit and prediction are 0.39 and 0.35, respectively (Fig 8). Only the relative size (\hat{A}) seemed to be associated with time although the correlation was relatively weak. According to Table 3 goodness of fit and prediction were 0.5 for \hat{A} as a function of all predictor variables, including time. Performing linear regression with \hat{A} as a function of time alone gave a correlation coefficient of $R^2 = 0.4$. The latter low correlation could have been caused by rotations of rhodoliths around an axis not perpendicular to the image plane between the different time points of the images (Fig 10).

The results of the pre-processing steps show that the illumination fluctuations could be reduced significantly. Furthermore, the reference plate enabled us to achieve color constancy and to correct camera zooming for all images. Although we were able to design a pre-processing pipeline to enhance the image quality and to compensate feature shifts caused by the experimental set up, some experimental steps can still be improved. For instance, in most of the excluded images, the color reference plate showed strong reflections due to a suboptimal material surface and the angle between camera and reference plate. Hence, the circle detection failed or the gamma correction produced wrong corrections if the reflection influenced too many circles. The strong reflections that appeared in the color reference plate may be avoided by selecting a material that gives minimal reflections. Also using an imaging setup where the location and angle of the reference plate are fixed and optimized to avoid strong reflections will improve the image quality. Furthermore using a camera with manual settings enables the use of a fixed white balance that reduces possible fluctuations in color. The necessity to apply illumination correction could maybe be avoided using an optimized stable light setup system with artificial light sources only. Reference plate and sample should be illuminated evenly throughout the whole image. By experience even in laboratory studies these conditions are often hard to achieve throughout the whole experiment. The presented semi-automated software solution can be used to evaluate the images even if not recorded under the optimized conditions, as shown here.

At this state the whole automated segmentation progress takes about 4 minutes for each image, where most of the time is spent on the pre-processing stage. As shown by [31] a significant saving of processing time is achievable by code optimizations and the use general-purpose

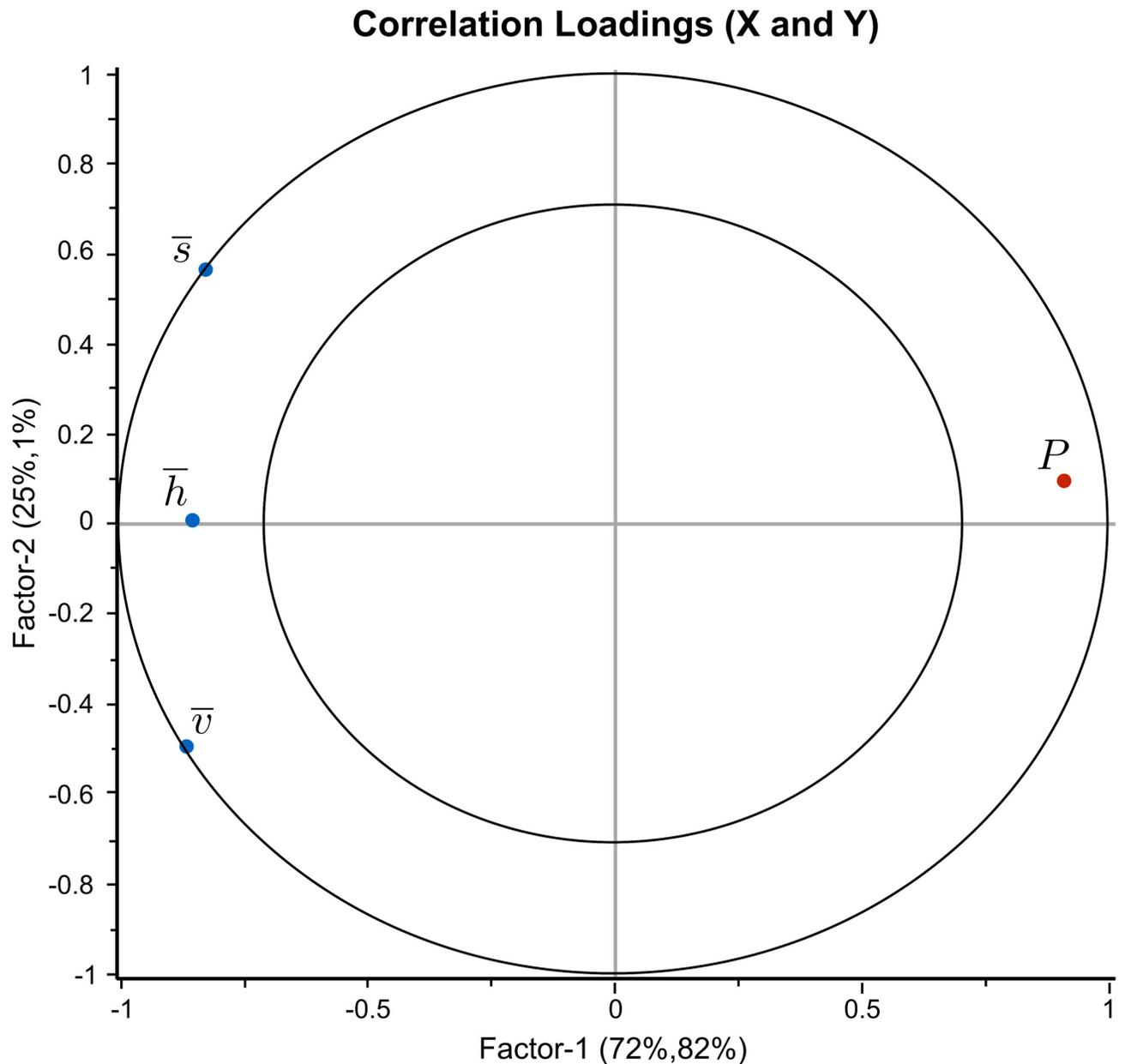


Fig 9. The correlation loading plot for the photosynthetic efficiency (P). It was obtained after PLS-regression with the HSV-variables (\bar{h} , \bar{s} , \bar{v}) as predictors (blue dots), and photosynthetic efficiency (P) as response (red dot). The outer circle represents 100% explained variance and the inner circle 50% explained variance for the different variables. The plot showed that the predictor variables are highly correlated to the response variable. A prediction of the photosynthetic efficiency using the HSV-values is therefore possible for our experiment.

doi:10.1371/journal.pone.0157329.g009

computing on graphics processing units (GPGPU). A similar acceleration can be expected for our approach, as no code optimization and no GPGPU has been applied for this study.

In comparison to this laboratory study, the analysis of underwater images recorded in the field is even more challenging for the issues discussed above as the levels of the inherent optical properties (IOP) will affect the image quality. Various concentrations of phytoplankton (Chl a), colored dissolved organic matter (cDOM) and total suspended matter (TSM) will alter the

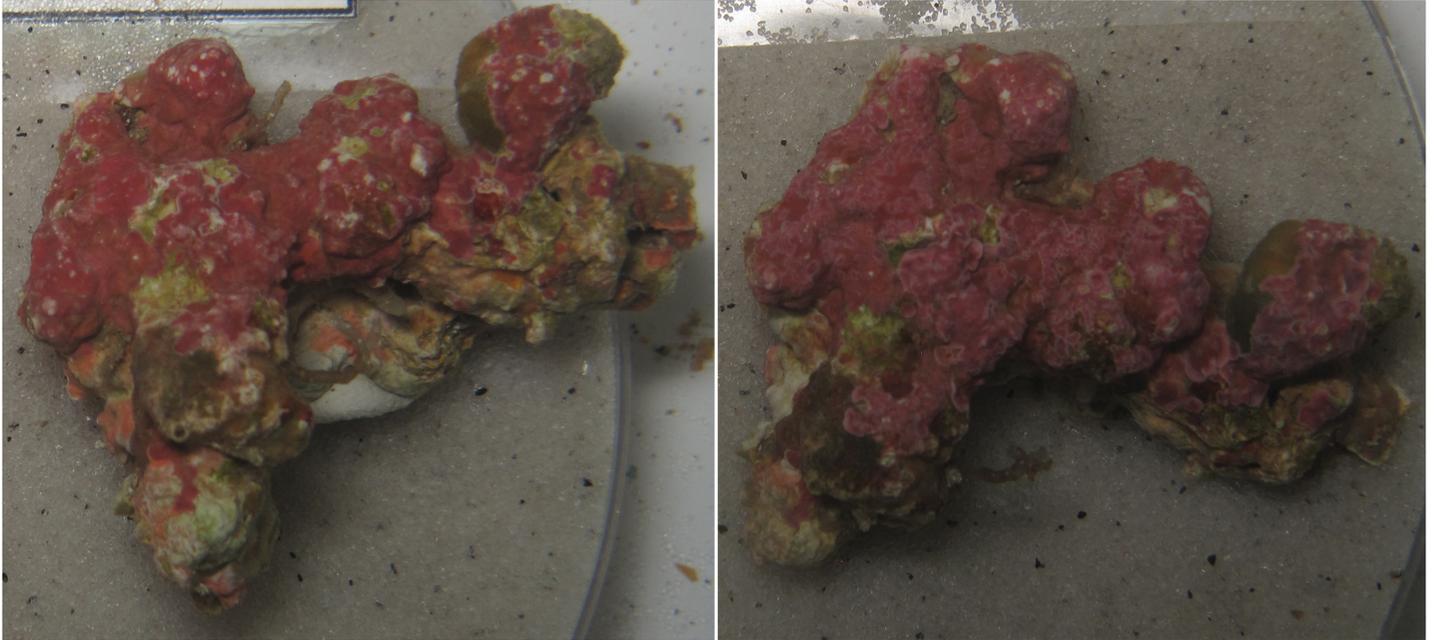


Fig 10. The relative position of the sample can have an effect on the segmentation result. Both images are displaying the same calcareous algae sample, but were recorded at different time points during the experiment. The calcareous algae sample was rotated around an axis not perpendicular to the image plane.

doi:10.1371/journal.pone.0157329.g010

contrast, sharpness and colors of objects due to absorption and/or scattering of light [42, 43]. In shallow waters, (0–20 m depth) reflection from the seafloor and varying light conditions related to the cloudiness and flickering effects (focusing/defocusing due to wave effects) would add even more challenges that should be considered [44]. Therefore, it must be expected that image (pre-)processing will be needed when operating under field conditions. As both the calcareous algae and sediments at the Peregrino field moves quite extensively (unpublished results), the monitoring should focus on measurements of possible color change of the segmented (visual) calcareous algae present in the image over time.

Furthermore, we believe that this methodological approach has the potential to be applied to studies with other species where color change is a stress indicator, e.g. tropical corals [7]. However, as individual species have species specific features (color, structure, shape, etc.) and their surroundings differ, species specific adjustment for computational analysis may be required.

Conclusion

This pilot study has shown that impact of sedimentation on calcareous algae samples can be detected by the presented computational approach for automatic size and color measurement from photos.

In addition to the previously published correlation between photosynthetic efficiency and sediment coverage [6], the present multivariate regression showed correlations between time and calcareous algae size, as well as correlations between fluorescence and calcareous algae color. Furthermore, we have shown that the use of an unsupervised learning algorithm is a promising attempt to create a classifier out of a sparsely labeled dataset. The chosen approach of an automatic assignment of the H²SOM prototypes was successful. The results of this pilot

study will enable researchers to conduct more studies in the future with different parameterizations and samplings to improve calcareous algae stress models.

The observed correlation between photosynthetic efficiency and the HSV-values is highly interesting for field studies, since photosynthetic efficiency measurements are impractical to perform manually in deeper waters. Color measurement in images has therefore the potential to play an important role in future field calcareous algae monitoring systems.

Supporting Information

S1 Text.

(PDF)

S2 Text.

(PDF)

S3 Text.

(PDF)

S1 Table.

(CSV)

Acknowledgments

We are grateful to the Chico Mendes Institute for Biodiversity Conservation for permitting the collection of the samples. The Instituto de Pesquisas Almirante Paulo Moreira offered their laboratory facilities to run the mesocosm trial with the support of Ricardo Coutinho, Carlos Gustavo Ferreira and Fernanda Siviero.

Author Contributions

Conceived and designed the experiments: JO IN IE MAOF FTST TWN. Performed the experiments: JO IN IE MAOF FTST TWN. Analyzed the data: JO TWN IE IN. Contributed reagents/materials/analysis tools: JO IN IE MAOF FTST TWN. Wrote the paper: JO IN IE MAOF FTST TWN.

References

1. Nilssen I, Ødegård Oy, Sørensen AJ, Johnsen G, Moline MA, Berge Jr. Integrated environmental mapping and monitoring, a methodological approach to optimise knowledge gathering and sampling strategy. *Marine Pollution Bulletin*. 2015; 96(1–2):374–383. Available from: <http://www.sciencedirect.com/science/article/pii/S0025326X15002477>. doi: [10.1016/j.marpolbul.2015.04.045](https://doi.org/10.1016/j.marpolbul.2015.04.045) PMID: [25956441](https://pubmed.ncbi.nlm.nih.gov/25956441/)
2. Buhl-Mortensen PI, Tenningen E, Tysseland ABS. Effects of water flow and drilling waste exposure on polyp behaviour in *Lophelia pertusa*. *Marine Biology Research*. 2015; 11(7):1–13. Available from: <http://www.tandfonline.com/doi/abs/10.1080/17451000.2014.993651>. doi: [10.1080/17451000.2014.993651](https://doi.org/10.1080/17451000.2014.993651)
3. Purser A. A Time Series Study of *Lophelia pertusa* and Reef Megafauna Responses to Drill Cuttings Exposure on the Norwegian Margin. *Plos One*. 2015; 10(7):e0134076. Available from: <http://dx.plos.org/10.1371/journal.pone.0134076> PMID: [26218658](https://pubmed.ncbi.nlm.nih.gov/26218658/)
4. Purser a, Ontrup J, Schoening T, Thomsen L, Tong R, Unnithan V, et al. Microhabitat and shrimp abundance within a Norwegian cold-water coral ecosystem. *Biogeosciences*. 2013; 10(9):5779–5791. doi: [10.5194/bg-10-5779-2013](https://doi.org/10.5194/bg-10-5779-2013)
5. Villas-Bôas AB, Tâmega FTS, Andrade M, Coutinho R, Figueiredo MAO. Experimental Effects of Sediment Burial and Light Attenuation on Two Coralline Algae of a Deep Water Rhodolith Bed in Rio de Janeiro, Brazil. *Cryptogamie, Algologie*. 2014; 35(1):67–76. Available from: <http://www.bioone.org/doi/abs/10.7872/crya.v35.iss1.2014.67>. doi: [10.7872/crya.v35.iss1.2014.67](https://doi.org/10.7872/crya.v35.iss1.2014.67)

6. Figueiredo MAO, Eide I, Reynier M, Villas-Bôas AB, Tâmega FTS, Ferreira CG, et al. The effect of sediment mimicking drill cuttings on deep water rhodoliths in a flow-through system: Experimental work and modeling. *Marine Pollution Bulletin*. 2015; 95(1):81–88. Available from: <http://www.sciencedirect.com/science/article/pii/S0025326X15002428>. doi: [10.1016/j.marpolbul.2015.04.040](https://doi.org/10.1016/j.marpolbul.2015.04.040) PMID: [25935812](https://pubmed.ncbi.nlm.nih.gov/25935812/)
7. Siebeck UE, Marshall NJ, Klüter a, Hoegh-Guldberg O. Monitoring coral bleaching using a colour reference card. *Coral Reefs*. 2006; 25(3):453–460. doi: [10.1007/s00338-006-0123-8](https://doi.org/10.1007/s00338-006-0123-8)
8. Hall-Spencer JM. Conservation issues relating to maerl beds as habitats for molluscs. *Journal of Conchology*. 1998; 2(2):271–286.
9. Steller DL, Riosmena-rodríguez R, Foster MS. 6 Living Rhodolith Bed Ecosystems in the Gulf of California. *Ecosystems*. 2001;(1944:):72–82.
10. Riosmena-Rodríguez R, Steller D, Hinojosa-Arango G, Foster M. Reefs that rock and roll: Biology and conservation of rhodolith beds in the Gulf of California. *The Gulf of California: Biodiversity and Conservation*. 2010;p. 49–71.
11. Tâmega FTS, Oliveira PS, Figueiredo MAO. Catalogue of the Benthic Marine Life for Peregrino Oil Field. Campos Basin, Brazil Instituto Biodiversidade Marinha, Rio de Janeiro. 2013;p. 140.
12. Basso D. Deep rhodolith distribution in the Pontian Islands, Italy: A model for the paleoecology of a temperate sea. *Palaeogeography, Palaeoclimatology, Palaeoecology*. 1998; 137(1–2):173–187. doi: [10.1016/S0031-0182\(97\)00099-0](https://doi.org/10.1016/S0031-0182(97)00099-0)
13. Henriques MC, Riosmena-Rodríguez R, Coutinho L, Figueiredo MAO Lithophylloideae and Mastophoroideae (Corallinales, Rhodophyta) from the Brazilian continental shelf. *Phytotaxa*. 2014; 190(1):112–129. Available from: <http://biotaxa.org/Phytotaxa/article/view/phytotaxa.190.1.9>. doi: [10.11646/phytotaxa.190.1.9](https://doi.org/10.11646/phytotaxa.190.1.9)
14. Littler MM, Littler DS, Hanisak MD. Deep-water rhodolith distribution, productivity, and growth history at sites of formation and subsequent degradation. *Journal of Experimental Marine Biology and Ecology*. 1991; 150(2):163–182. Available from: <http://www.sciencedirect.com/science/article/pii/S0022098191900666>. doi: [10.1016/0022-0981\(91\)90066-6](https://doi.org/10.1016/0022-0981(91)90066-6)
15. Littler MM, Littler DS, Blair SM, Norris JN. Deep-water plant communities from an uncharted seamount off San Salvador Island, Bahamas: distribution, abundance, and primary productivity. *Deep Sea Research Part A Oceanographic Research Papers*. 1986; 33(7):881–892. Available from: <http://www.sciencedirect.com/science/article/pii/0198014986900038>. doi: [10.1016/0198-0149\(86\)90003-8](https://doi.org/10.1016/0198-0149(86)90003-8)
16. Kempf M. Notes on the benthic bionomy of the N-NE Brazilian shelf. *Marine Biology*. 1970; 5(3):213–224. doi: [10.1007/BF00346909](https://doi.org/10.1007/BF00346909)
17. Foster MS. Rhodoliths: Between rocks and soft places. *Journal of Phycology*. 2001; 37(5):659–667. doi: [10.1046/j.1529-8817.2001.00195.x](https://doi.org/10.1046/j.1529-8817.2001.00195.x)
18. Amado-Filho GM, Moura RL, Bastos AC, Salgado LT, Sumida PY, Guth AZ, et al. Rhodolith beds are major CaCO₃ BIO-factories in the tropical south West Atlantic. *PLoS ONE*. 2012; 7(4):e35171. doi: [10.1371/journal.pone.0035171](https://doi.org/10.1371/journal.pone.0035171) PMID: [22536356](https://pubmed.ncbi.nlm.nih.gov/22536356/)
19. Nelson W. Calcified macroalgae—critical to coastal ecosystems and vulnerable to change: a review. *Marine and Freshwater Research*. 2009; 60(8):787–801. Available from: <http://www.publish.csiro.au/?paper=MF08335>. doi: [10.1071/MF08335](https://doi.org/10.1071/MF08335)
20. Davies AJ, Roberts JM, Hall-Spencer J. Preserving deep-sea natural heritage: Emerging issues in offshore conservation and management. *Biological Conservation*. 2007; 138(3–4):299–312. doi: [10.1016/j.biocon.2007.05.011](https://doi.org/10.1016/j.biocon.2007.05.011)
21. Schoening T, Osterloff J, Nattkemper TW. RecoMIA—Recommendations for Marine Image Annotation: Lessons Learned and Future Directions. *Frontiers in Marine Science*. 2016; 3(59). Available from: http://www.frontiersin.org/deep-sea_environments_and_ecology/10.3389/fmars.2016.00059/abstract.
22. Kohler KE, Gill SM. 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 and Geosciences*. 2006; 32(9):1259–1269. doi: [10.1016/j.cageo.2005.11.009](https://doi.org/10.1016/j.cageo.2005.11.009)
23. Ontrup J, Ehnert N, Bergmann M, Nattkemper TW. Biigle—Web 2.0 enabled labelling and exploring of images from the Arctic deep-sea observatory HAUSGARTEN. In: *OCEANS'09 IEEE Bremen: Balancing Technology with Future Needs*. IEEE; 2009. p. 1–7.
24. Purser A, Bergmann M, Lundälv T, Ontrup J, Nattkemper TW. Use of machine-learning algorithms for the automated detection of cold-water coral habitats: A pilot study. *Marine Ecology Progress Series*. 2009; 397:241–251. doi: [10.3354/meps08154](https://doi.org/10.3354/meps08154)
25. Bergmann M, Langwald N, Ontrup J, Soltwedel T, Schewe I, Klages M, et al. Megafaunal assemblages from two shelf stations west of Svalbard. *Marine Biology Research*. 2011; 7(6):525–539. doi: [10.1080/17451000.2010.535834](https://doi.org/10.1080/17451000.2010.535834)

26. Schoening T, Bergmann M, Ontrup J, Taylor J, Dannheim J, Gutt J, et al. Semi-automated image analysis for the assessment of megafaunal densities at the Arctic deep-sea observatory HAUSGARTEN. *PLoS ONE*. 2012; 7(6):e38179. doi: [10.1371/journal.pone.0038179](https://doi.org/10.1371/journal.pone.0038179) PMID: [22719868](https://pubmed.ncbi.nlm.nih.gov/22719868/)
27. Bergmann M, Klages M. Increase of litter at the Arctic deep-sea observatory HAUSGARTEN. *Marine Pollution Bulletin*. 2012; 64(12):2734–2741. doi: [10.1016/j.marpolbul.2012.09.018](https://doi.org/10.1016/j.marpolbul.2012.09.018) PMID: [23083926](https://pubmed.ncbi.nlm.nih.gov/23083926/)
28. Osterloff J, Nilssen I, Nattkemper TW. A computer vision approach for monitoring the spatial and temporal shrimp distribution at the LoVe observatory. *Methods in Oceanography*. 2016;p. –. Available from: <http://www.sciencedirect.com/science/article/pii/S2211122015300062>. doi: [10.1016/j.mio.2016.03.002](https://doi.org/10.1016/j.mio.2016.03.002)
29. Beijbom O, Edmunds PJ, Roelfsema C, Smith J, Kline DI, Neal BP, et al. Towards Automated Annotation of Benthic Survey Images: Variability of Human Experts and Operational Modes of Automation. *PLoS ONE*. 2015; 10(7):e0130312. doi: [10.1371/journal.pone.0130312](https://doi.org/10.1371/journal.pone.0130312) PMID: [26154157](https://pubmed.ncbi.nlm.nih.gov/26154157/)
30. Ontrup J, Ritter H. Hyperbolic Self-Organizing Maps for Semantic Navigation. *Advances in Neural Information Processing Systems* 14. 2002; 14(14):1417–1424.
31. Schoening T, Langenkämper D, Steinbrink B, Brün D, Nattkemper TW. Rapid image processing and classification in underwater exploration using advanced high performance computing. In: *OCEANS 2015—MTS/IEEE Washington*; 2015. p. 1–5.
32. Schoening T, Kuhn T, Jones DOB, Simon-Lledo E, Nattkemper TW. Fully automated image segmentation for benthic resource assessment of poly-metallic nodules. *Methods in Oceanography*. 2016;p. –. Available from: <http://www.sciencedirect.com/science/article/pii/S2211122015300104>. doi: [10.1016/j.mio.2016.04.002](https://doi.org/10.1016/j.mio.2016.04.002)
33. Box GEP, Hunter WG, Hunter JS. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. Wiley Series in Probability and Mathematical Statistics. 1978; 1st:653.
34. Foster DH. *Color constancy*. vol. 51 of The Wiley-IS&T Series in Imaging Science and Technology. Wiley; 2010.
35. Stokes, M, Anderson, M, Chandrasekar, S, Motta, R. A standard default color space for the internet-srgb. In: *Microsoft and Hewlett-Packard Joint Report. Society for Imaging Science and Technology*; 1996. p. 1–17. Available from: http://fp.optics.arizona.edu/opti588/reading/Stokes_sRGB_colorspace_detail.pdf.
36. Yuen H, Princen J, Illingworth J, Kittler J. Comparative study of Hough Transform methods for circle finding. *Image and Vision Computing*. 1990; 8(1):71–77. doi: [10.1016/0262-8856\(90\)90059-E](https://doi.org/10.1016/0262-8856(90)90059-E)
37. Gonzalez RC, Woods RE. *Digital Image Processing*. 3rd ed. Upper Saddle River, NJ: Pearson Prentice Hall; 2007.
38. Wold S, Martens H, Wold H. The multivariate calibration problem in chemistry solved by the PLS method. In: *Matrix Pencils*. vol. 973. Springer; 2012. p. 286–293. Available from: <http://dx.doi.org/10.1007%2fBFb0062108>.
39. Forina M, Lanteri S, Casale M. *Multivariate calibration*. vol. 1158. John Wiley & Sons; 2007.
40. Wold S. Cross-Validatory Estimation of the Number of in Factor and Principal Components Components Models. *Technometrics*. 1978; 20(4):397–405. doi: [10.1080/00401706.1978.10489693](https://doi.org/10.1080/00401706.1978.10489693)
41. Esbensen K. *Multivariate Data Analysis—In Practice: An Introduction to Multivariate Data Analysis and Experimental Design*. 5th ed. Camo Process AS; 2002.
42. Iocccg. *Remote Sensing of Ocean Colour in Coastal, and Other Optically-Complex Waters*. In: *Reports of the International Ocean Colour Coordinating Group 3*; 2000. p. 144.
43. Johnsen G, Volent Z, Sakshaug E, Sigemes F, Pettersson LH. 6. Remote sensing in the Barents Sea. In: *Sakshaug E, Johnsen GH, Kovacs KM, editors. Ecosystem Barents Sea*. Tapir Academic Press; 2009. p. 139–166.
44. Shihavuddin A, Gracias N, Garcia R. Online Sunflicker Removal using Dynamic Texture Prediction. In: *VISAPP (1)*; 2012. p. 161–167.