SoftwareX 6 (2017) 69–80

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

PAPARA(ZZ)I: An open-source software interface for annotating photographs of the deep-sea

Yann Marcon[a,b,*], Autun Purser[a]

[a] *Deep Sea Ecology and Technology, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany*

[b] *MARUM Center for Marine Environmental Sciences, Bremen, Germany*

ARTICLE INFO

Article history:
Received 15 January 2016
Received in revised form
20 January 2017
Accepted 8 February 2017

Keywords:
PAPARA(ZZ)I
Image annotation
WoRMS
Benthic
Megafauna
Density
Distribution
Composition
Abundance
Richness

ABSTRACT

PAPARA(ZZ)I is a lightweight and intuitive image annotation program developed for the study of benthic megafauna. It offers functionalities such as free, grid and random point annotation. Annotations may be made following existing classification schemes for marine biota and substrata or with the use of user defined, customised lists of keywords, which broadens the range of potential application of the software to other types of studies (e.g. marine litter distribution assessment). If Internet access is available, *PAPARA(ZZ)I* can also query and use standardised taxa names directly from the World Register of Marine Species (WoRMS). Program outputs include abundances, densities and size calculations per keyword (e.g. per taxon). These results are written into text files that can be imported into spreadsheet programs for further analyses. *PAPARA(ZZ)I* is open-source and is available at http://papara-zz-i.github.io. Compiled versions exist for most 64-bit operating systems: Windows, Mac OS X and Linux.

Code metadata

Current code version	*PAPARA(ZZ)I v2.6*
Permanent link to code/repository used of this code version	https://github.com/ElsevierSoftwareX/SOFTX-D-16-00015
Legal Code License	*GNU General Public License* (http://www.gnu.org/copyleft/gpl.html)
Code versioning system used	*Github version control*
Software code languages, tools, and services used	*MATLAB, Java*
Compilation requirements, operating environments & dependencies	*MATLAB 2010a-2015b, MATLAB Compiler, Optional: MATLAB Image Processing Toolbox*
If available Link to developer documentation/manual	https://github.com/PAPARA-ZZ-I/PAPARA-ZZ-I/blob/master/PAPARAZZI_2.6_UserManual.pdf
Support email for questions	Yann.Marcon@awi.de

Software metadata

Current software version	*PAPARA(ZZ)I v2.6*
Permanent link to executables of this version	https://papara-zz-i.github.io/ *or* https://github.com/PAPARA-ZZ-I/PAPARA-ZZ-I/tree/master/PAPARA(ZZ)I_installers
Legal Software License	*GNU General Public License* (http://www.gnu.org/copyleft/gpl.html)
Computing platforms/Operating Systems	*64-bit Windows, Mac OS X, Linux*
Installation requirements & dependencies	*MATLAB Runtime v9.0 (included in the PAPARA(ZZ)I installers)*
If available, link to user manual - if formally published include a reference to the publication in the reference list	https://github.com/PAPARA-ZZ-I/PAPARA-ZZ-I/blob/master/PAPARAZZI_2.6_UserManual.pdf
Support email for questions	Yann.Marcon@awi.de

* Corresponding author at: Deep Sea Ecology and Technology, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany.
E-mail address: Yann.Marcon@awi.de (Y. Marcon).

http://dx.doi.org/10.1016/j.softx.2017.02.002

1. Motivation and significance

Economic interest in the deep-sea has been growing in the last decades due to the increasing stress and demand on onshore resources. The deep-sea is known to host large quantities of ore (sulphide deposits, polymetallic nodules, etc.) [1], which are becoming economically viable for exploitation. However, very few regions of the deep sea have been visually investigated and information about the deep seafloor is very scarce due to its remoteness and difficulty of access [2,3]. The lack of information on deep-sea fauna limits our understanding of the largest ecosystem on Earth. This is especially true in the current economic climate, where deep-sea mining activities are foreseen in the near future, as well as in regions where deep-sea fishing is known to impact on previously unexploited deep-sea fauna [4]. The current knowledge on how much these activities will impact the ecosystem is limited. The lack of data and knowledge of ecosystem processes in these regions is recognised as a cause of concern within the scientific community. To address this, many recent studies have highlighted the need for marine policy and management strategies to protect deep-sea ecosystems [5–10]. However, we still lack comprehensive understanding of these deep-sea ecosystems, including data on baseline ecosystem conditions prior to any human disturbance and significance of other anthropogenic effects. For instance, information about the abundance, richness, composition and spatial distribution of the deep-sea fauna is only available for small, discrete localised areas of the seafloor.

Optical surveys are suitable for studying the baseline conditions of the larger organisms (megafauna and some macrofauna) without disturbing the seafloor [11]. Not only are these surveys less invasive than traditional fauna and sediment sampling of the seafloor with box cores, trawls etc., but they may also sample larger areas of seafloor to give a better representation of the ecosystem at a larger scale. However, due to low visibility in the deep sea, hundreds to thousands of images need to be collected, in order to make statistically sound studies on the benthic and benthopelagic fauna. Analysing such large numbers of images requires considerable efforts that can be mitigated with the use of specialised image annotation software.

Several existing web-based programs for annotating and labelling marine imagery offer various ranges of functionalities to the user [12–17]. As the number of web-based technologies for scientific research increases, the general workflow is not always efficient or easy to use. For instance, some solutions allow users to analyse images online and to store their annotations in a central database (e.g. CATAMI, SQUIDLE, BIIGLE and DIAS). However, these web-based image analysis tools do not allow users to upload their own datasets. Users are required to contact the web-administrator or host institute, in order to upload data on the website. This introduces delays before the analysis of the data can start and also depends on the availability of the administrator responsible for the data upload. Additional delays are caused by the long loading times of high-resolution images from web-based tools, which is a hurdle when analysing large numbers of images. Finally, uploading datasets to a host server is not always possible in the field (particularly during offshore marine field work where internet access is either limited or not available) or due to copyright limitations on the images. By contrast, offline tools, such as *PAPARA(ZZ)I*, enable users to analyse images immediately, in any location, and independent of Internet availability whilst also keeping full control of their dataset (i.e. in case of copyright concerns or industrial sensitivity).

In this work we present *PAPARA(ZZ)I*, an open-source and lightweight (simple and fast workflow) application for point annotating images that are taken perpendicular to an assumed flat seabed. It has a Graphical User Interface (GUI) and a streamlined workflow designed to make the analysis of high numbers of images fast and efficient (e.g. annotations require few mouse clicks, settings are not reset for every image, all functionalities can be called up from the main GUI without jumping through menus). *PAPARA(ZZ)I* shares similar functionalities with other offline image analysis programs that are also available for free (e.g. VARS [12], CPCe [13], NICAMS [17], FISH_ROCK [18], ImageJ [19]). However these programs do not offer the same degree of practicality or flexibility when analysing a large number of images with point annotations. For instance, VARS requires an online connection for querying the knowledgebase from the Monterey Bay Aquarium Research Institute's (MBARI) during the analysis, or that a personal PostgreSQL database server be setup locally, which requires a high level of computer-user proficiency. ImageJ requires that users load each image individually, redefine settings in various menus for every image (e.g. scale length), and keep a log of all annotations. These steps make the workflow cumbersome and less efficient, especially when analysis of thousands of images is required. CPCe only allows the analysis of randomly selected points (i.e. no free points or point grid analysis) and only runs on Windows platforms. FISH_ROCK, NICAMS and VARS do not allow random or gridded point counts (only free points mode).

None of the existing programs allows the user to (1) randomise the image order (except VARS), (2) set a limit to the maximum display scale during the annotations, and (3) filter annotations while browsing through a dataset. These functionalities are important for a marine ecologist, to reduce analysis biases (Sections 2.2.4 and 2.2.5), as well as to retrieve a particular annotated feature or to fine-tune the analysis of specific taxa (Section 2.2.9). The *PAPARA(ZZ)I* annotating regime was developed to address these specific needs. It is intuitive, requires few mouse clicks, settings can be saved and maintained throughout a dataset, and all functionalities can be called from the main GUI without having to scroll through menus. *PAPARA(ZZ)I* is simple to install and runs on Windows, Mac and Linux. An installer is provided for all three operating systems.

PAPARA(ZZ)I allows users to annotate images offline using free points, point grid or random point arrays with either existing classification schemes [20–23] or customised, user defined lists of keywords. When Internet connectivity is available, users have the extra option to query the World Register of Marine Species (WoRMS) [24] directly from the main GUI of the program and use these query results to annotate the images. *PAPARA(ZZ)I* provides a range of functions: (1) image scaling, (2) size measurements of objects (lengths and widths), (3) randomisation of image order to prevent analysis bias caused by increasing user habituation to the dataset during analysis, (4) maximum display scale control to reduce inter-observer bias or prevent biases when a user changes monitor size during analysis, (5) definition of the useable image area, (6) lens distortion compensation (for area and size calculations only), (7) annotation filter for quick retrieval of annotated features, (8) batch correction of annotations, (9) automatic transcription of all annotations and measurements into text files, and (10) direct export of per-taxon abundances, densities, and size distribution data to summary text files. Each of these features are described in detail in the following sections where we also outline key features that are unique to *PAPARA(ZZ)I*.

This work presents the functionalities of the *PAPARA(ZZ)I* image annotation program and its interest to marine ecologists. *PAPARA(ZZ)I* projects have a simple and clear data structure based on the use of text-files. Output files are tab-delimited text files that can be read and manipulated with other programs for further analyses (text editors, spreadsheets, R, PRIMER, SPSS, etc.). It is an open-source software released under the GNU General Public License v3 [25]. Therefore, the code can be modified as needed to meet the needs of every user. The program is provided with a user manual, which includes additional instructions for installing *PAPARA(ZZ)I* on each operating system, as well as screenshots and detailed how-to explanations for all functionalities presented in this work. *PAPARA(ZZ)I* was developed with MATLAB and can be run directly from the MATLAB command window (by calling the *PAPARAZZI.m* script). However, if the user does not have a MATLAB licence, it can also be installed as a standalone application on most 64-bit operating systems (Windows, Mac OS X, Linux) by running the appropriate installer files (exact commands to run the installer files are provided in the user manual). The source code, the installer files and the user manual are available at http://papara-zz-i.github.io. A test image dataset can also be downloaded, which contains examples for every input file and allows users to test all the functionalities of the program.

2. Software description

PAPARA(ZZ)I was initially developed to answer an immediate need for a reliable and easy-to-use program for rapid annotation of images. To this end, the workflow was purposely streamlined with the number of steps involved in the analysis kept as few as possible. Once *PAPARA(ZZ)I* is installed, the workflow is very straightforward: (1) copying all images into a single folder (referred to as the *project folder*), (2) starting *PAPARA(ZZ)I* and opening the project folder, (3) selecting a classification scheme or connecting to WoRMS [24], (4) annotating the images, and (5) exporting the results. A few optional steps can also be added to this base workflow, such as randomising the image order, generating regularly or randomly distributed points for classification across the images, defining the image useable area, limiting the maximum display scale (maximum zoom), or importing the camera and lens distortion (6th order radial + tangential distortion) parameters. These functionalities are described in Section 2.2.

2.1. Software architecture

PAPARA(ZZ)I was developed in the MATLAB language with the addition of some Java components in the graphical user interface (GUI). The program works as a standalone application in 64-bit Windows, Mac OS X and Linux operating systems, and only requires the MATLAB runtime installation which is provided with the *PAPARA(ZZ)I* package. Alternatively, *PAPARA(ZZ)I* can be used by running the source code directly from MATLAB. The code was written using MATLAB R2015b, and is fully operational in older versions (tested using MATLAB R2010a). Extra MATLAB toolboxes are not required to run the program.

PAPARA(ZZ)I can work with images that are stored either locally on the computer or on a work server (Fig. 1). The latter option has the advantage of allowing several users to work on the same project at the same time provided that they use distinct usernames and that they have write access to the server. Raw annotations (i.e. keywords and pixel position of annotation within the image) are directly written into text files that are saved in subdirectories of the open project folder (Fig. 1). The annotations are stored in a different text file for each image/user combination and only one text file is opened with write-access at a time. This ensures that each annotation is automatically saved as it is marked, limiting data loss or corruption should the program computer or server connection (if using a work server), unexpectedly stop working. Existing annotations are automatically loaded by *PAPARA(ZZ)I* when re-opening the project folder.

Access to the Internet is only required to be able to access the web-services of the World Register of Marine Species [24] but it is not a compulsory requirement in order to use *PAPARA(ZZ)I*. The opportunity to be able to work offline is indeed very important for deep-sea scientists during research expeditions and the restricted Internet access commonly found on research vessels was the main motivation for the *PAPARA(ZZ)I* project [26].

When the image analysis is finished, the output annotation and result files may be uploaded by the user onto long-term data centres or repositories [27–33] Scientists can then download the annotation text files and read them with *PAPARA(ZZ)I* on any computer. Furthermore, the output text files are all written in formats that can be easily read and manipulated by other programs or scripts to maximise the field of possibilities for analysis of the annotation results (Fig. 2). For instance, *PAPARA(ZZ)I* annotations can likely be imported into databases, such as those associated with existing web-based annotation tools [12,14] with little data manipulation.

2.2. Software functionalities

PAPARA(ZZ)I was primarily designed for analysing benthic images from downward looking cameras. Therefore, the current stable version of the program (v2.6) includes functionalities that cover several types of point annotation modes (i.e. point grid and random points) that are commonly used in the fields of marine biology and ecology [16,14,34–36]. All functionalities and options are described in this section.

2.2.1. Annotation modes

PAPARA(ZZ)I allows images to be annotated in three different ways: (1) free point annotations, (2) regularly-distributed point (point grid) and (3) randomly-distributed point annotations. It is also possible to switch back and forth between these three types of analysis without losing any data. The annotations and the pixel coordinates of all the points and annotations are recorded automatically.

Free annotations are made by annotating features of interest directly on the images. This is done by first selecting the desired keyword (e.g. taxon name, substratum type, etc.) from the list in the left panel of the GUI and then by selecting the feature of interest on the image itself. *PAPARA(ZZ)I* immediately saves the annotations in a text file and displays a dot circle on the image where the annotation was made. The user can make as many annotations as necessary per image.

When using the other analysis modes – regularly or randomly distributed points – the number of annotation points per image is set by the user at the beginning of the analysis of the image set, and is maintained throughout the analysis of all images. The regularly or randomly distributed points are automatically generated by *PAPARA(ZZ)I* and are saved in the corresponding text file with the default "empty" attribute. The user can then change the attribute by selecting the desired point and choosing the appropriate keywords from the list. These modes allow the user to select several keywords for a same point in order to describe different observations (substratum, fauna, etc.). On the screen, points with the "empty" attribute are by default displayed with three coloured circles (white, blue, black) to ensure that no point is missed during the analysis. These three colours were chosen to stand out against low contrast areas of the image. Conversely, points with a valid attribute are displayed by a single blue circle, which allows the user to quickly distinguish annotated from unannotated points.

Techniques that are based on regularly and randomly distributed points are often used in ecological studies to guarantee a representation of the analysed area at a resolution determined by the number of points, as well as to limit user-induced bias by removing the ability of the user to label 'charismatic' features in preference to less immediately obvious features. These methods are predominantly used for studies that focus on highly productive

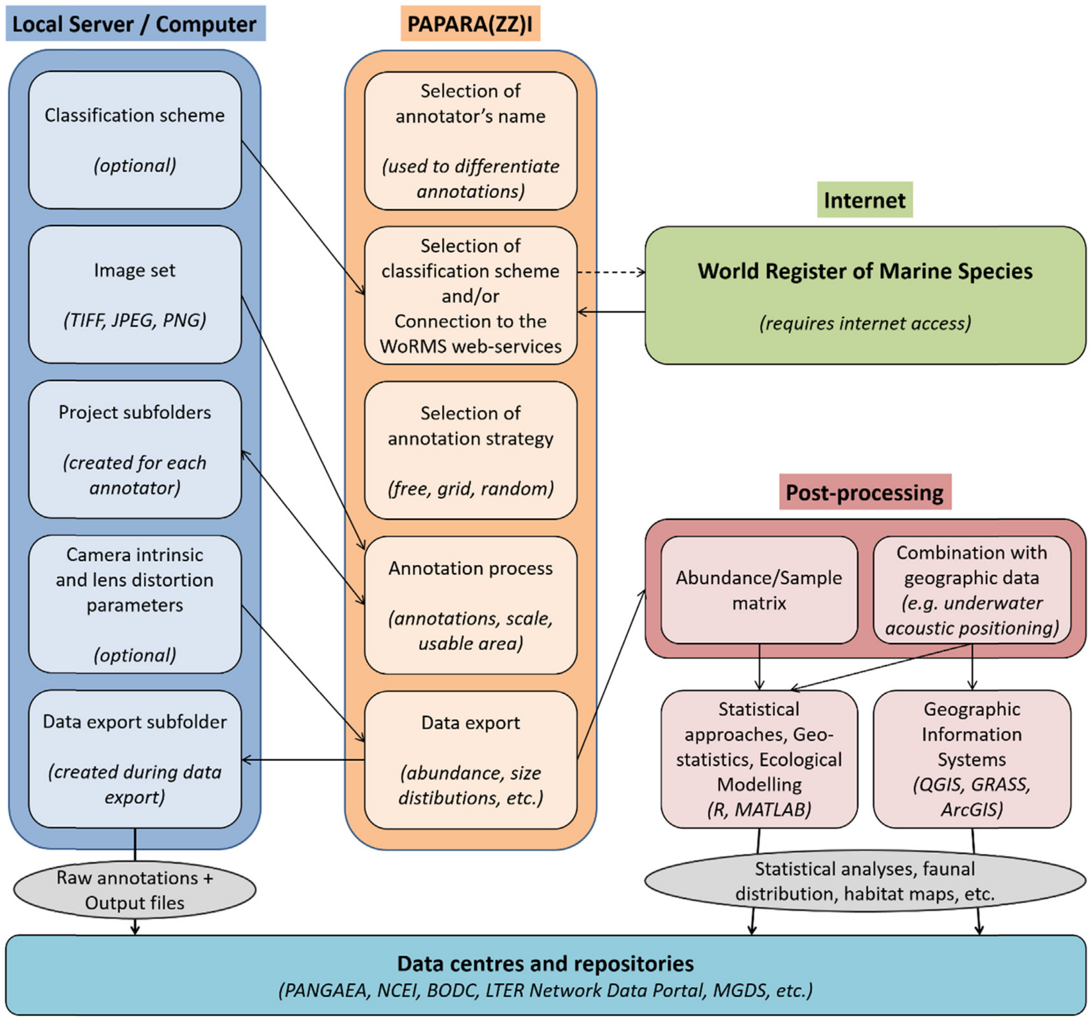

Fig. 1. A Architecture and workflow of *PAPARA(ZZ)I*. *PAPARA(ZZ)I* inputs and outputs can be located on a local computer or server. Connection to the WoRMS web-services is optional if Internet is available. Exported data can be used for further statistical analyses as well as for spatial representations of fauna distribution. PAPARA(ZZ)I output files and additional analysis results are intended to be uploaded to scientific data repositories to ensure transparency and reuse of the results.

shallow and coastal marine ecosystems where high faunal abundance and diversity occur, such as within coral reef ecosystems [16, 37–44]. In the deep sea, where megafauna distribution can be too scarce for these unbiased techniques to best represent the ecosystem, free counting techniques may be more appropriate, where all individuals are important members of the sparse community (i.e. in areas with low megafauna abundance) [45–55].

2.2.2. Image scale

Marine ecologists generally use laser pointers with a known spacing during the image acquisition in order to be able to scale their observations. Alternatively, an object of known size can also be placed in the field of view of the camera to determine the image scale. In both cases this information can be used in *PAPARA(ZZ)I* to assess the dimensions of the images and features of interest.

Users can draw a scale bar on the images and assign a length value in metres (e.g. the distance between laser points). *PAPARA(ZZ)I* is then able to convert lengths in image units (pixels) to lengths in world units (metres). This function is based on the assumption that the seabed in the image is flat and that the image was taken perpendicularly to it.

2.2.3. Size measurements

PAPARA(ZZ)I uses the scale information for the calculation of the image area in world units (i.e. square metres), and of faunal densities (ind/m^2) [56]. Additionally, it allows users to make size measurements on any annotated feature. Size measurements are done from the main GUI by selecting both ends of a feature of interest using the size measurement tool. The results of the size measurements can be found in the summary files that are generated during the data export.

The size measurements are *directional* in that they can be used to determine the direction angle of the sized feature in relation to the top side of the image. The direction of a size measurement is set when the measurement is made: the first point defines the front end (e.g. fish head) of the feature being measured and the second point defines the rear end (e.g. fish tail). *PAPARA(ZZ)I* defines the direction as the clockwise angle starting from the top of the image.

Direction angles are by-products of size measurements and they are automatically computed during the data export for every size measurement that is made. The resulting angles are intended to be combined with camera heading (or yaw) data in order to determine directions in relation to North (Fig. 3). Although we are unaware of any ecological study that has used such functionality, direction angles may find uses in future works, for instance to assess the connection between biota and orientation of geological or physical features on the seafloor.

2.2.4. Maximum display scale

The maximum display scale ensures that images can be displayed with the exact same size to the users, independent of

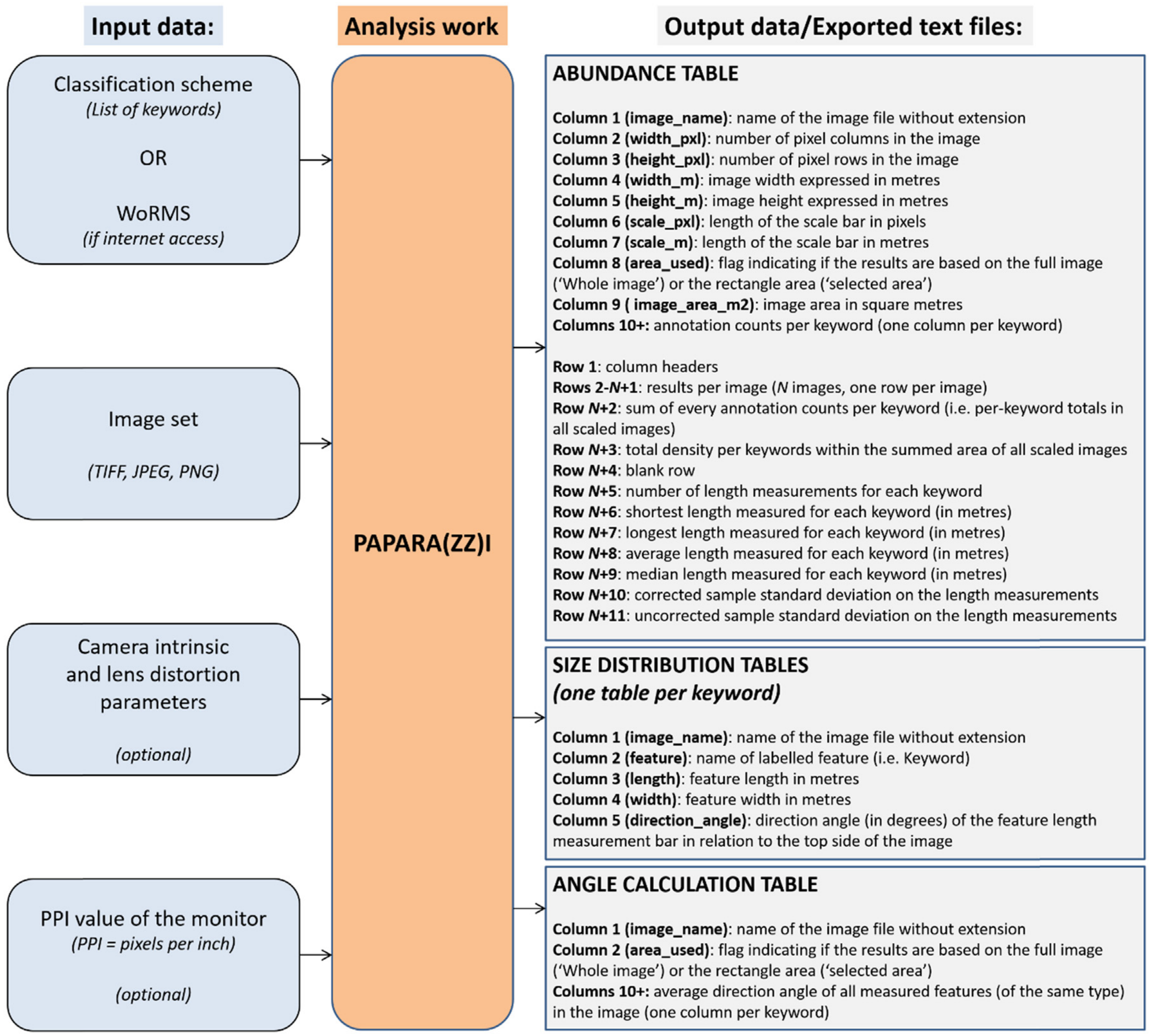

Fig. 2. Flow chart describing the PAPARA(ZZ)I inputs and outputs. Apart from the set of images to analyse, PAPARA(ZZ)I only requires a list of keywords to be provided. All other inputs are optional. Output data are stored as tables within text files with a simple structure. The structure of the output files is described. The structure of the output files is the same for all annotation modes (free points, point grid, random points).

what type of monitor is used for the analysis. A display scale ratio can be set directly on the main GUI. A scale ratio of 1:10 means that 1 cm on the screen represents 10 cm on the scaled image.

In some cases, setting a maximum display scale, i.e. a maximum level of zoom that users cannot exceed, helps reduce potential sources of analysis bias. Bias arises when different users analyse the same dataset [45,58,59]. For instance, in the case of marine ecological studies, users who tend to zoom in on images more than others will likely detect smaller organisms and find higher fauna abundances than users who go less into detail. A maximum display scale would prevent users from zooming past a certain level and ensure that all users analyse the dataset under the same conditions.

When setting the maximum display scale, users must also provide the resolution in *pixels per inch* (*PPI*) of the monitor that is used for the analysis. The monitor *PPI* value allows the program to determine the exact display size on the given monitor. If unknown, the monitor *PPI* is can be calculated with the following formula:

$$PPI = \frac{\sqrt{vpxl^2 + hpxl^2}}{D}.$$

Where the variables *hpxl* and *vpxl* are the horizontal and vertical resolutions of the monitor (in pixels), and *D* is the length of the monitor diagonal (in inches). The values of *hpxl*, *vpxl* and *D* can be found in the technical specifications of the monitor. The maximum display scale works only with images for which a scale bar has been defined.

2.2.5. Randomisation of image order

At the start of a new project, the program gives users the possibility to randomise the order of the images in the project folder. The purpose of the image randomisation is to reduce potential biases caused by the user habituation to the dataset. For the same user and dataset, the quality of the analysis may vary during the course of the work, as the user gets 'habituated' to the data. If the images are sorted chronologically or geographically, we expect that such habituation may introduce erroneous trends or false gradients in the results between the beginning and the end of the dataset. Randomising the image order may help reduce such bias.

If used, the randomisation process is done automatically at the start of the project. The random image order is saved in a text file located in the project folder and is reloaded automatically each time the same user reopens the project.

2.2.6. Lens distortion correction

Lens distortion is a type of optical aberration, which is caused by the geometry and misalignment of the camera optical components [60,61]. The most common types are symmetric radial

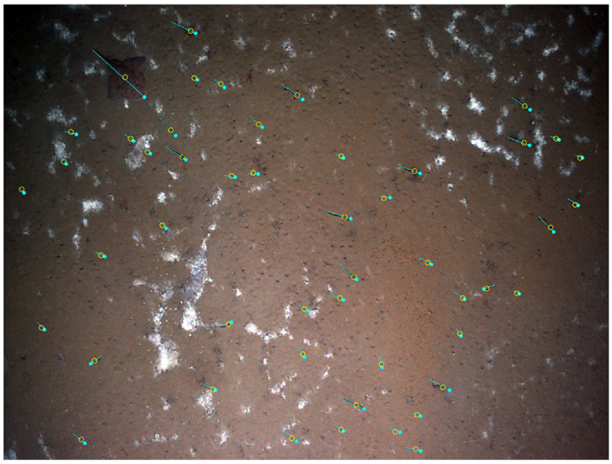

Fig. 3. Size measurements (blue lines) are directional: the dotted side indicates the front end of the annotated feature (yellow circle). Here, size measurements of Zoarcidae (*Lycodes squamiventer*) and Rajidae fishes can be used to estimate bottom current directions from images of the Haakon Mosby Mud Volcano. In this case, calculated mean angles per taxon (123° for Rajidae and 132° for Zoarcidae) combined with camera yaw angle (N81° W) suggest a bottom current direction between N42° E and N51° E (photo taken with WHOI AUV Sentry during cruise MSM16/2 [57]). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

distortion and decentring (or tangential) distortion. The symmetric radial distortion, related to the shape of the lens, causes the size of projected pixels (in world units) to change symmetrically with distance from the optical axis (e.g. barrel, pincushion or wave distortions) [62]. The tangential distortion, caused by the lens not being perfectly centred on the optical axis or parallel to the camera sensor, adds non-symmetric distortions to the images [60,63]. Either way, the main effect of lens distortion on an image is that the size of the projected pixels in world units is not uniform across the image area. This effect may be further amplified underwater by the water–air boundary if the lens that is used was not specifically designed for underwater use.

Therefore, depending on the lens/camera system that is used, and without adequate image correction, lens distortion may have a noticeable impact on the accuracy of the size measurements made on the images (e.g. fisheye lens). *PAPARA(ZZ)I* can compensate all size measurements and area calculations for lens distortion (Fig. 4), provided that the lens distortion coefficients are known. For the correction for lens distortion, *PAPARA(ZZ)I* uses the Brown–Conrady distortion model. It assumes a pinhole camera model and corrects for both radial and tangential distortions [60,62,63].

The radial distortion is defined by:

$$x_{distorted} = x + x\left(k_1 r^2 + k_2 r^4 + k_3 r^6\right)$$
$$y_{distorted} = y + y\left(k_1 r^2 + k_2 r^4 + k_3 r^6\right).$$

The tangential distortion is defined by:

$$x_{distorted} = x + \left[2\rho_1 xy + \rho_2\left(r^2 + 2x^2\right)\right]$$
$$y_{distorted} = y + \left[\rho_1\left(r^2 + 2y^2\right) + 2\rho_2 xy\right].$$

Where $r^2 = (x^2 + y^2)$ and (x, y) describe the pixel location of a point in the undistorted image, k_1, k_2, k_3 are the radial distortion coefficients, and ρ_1 and ρ_2 are the tangential distortion coefficients.

The correction for lens distortion is automatically applied during the data export if the camera intrinsic parameters and lens distortion coefficients were provided by the user. Camera intrinsic parameters include the pixel location of the optical centre, the focal length, the size of the camera sensor in pixels and millimetres, and the skew coefficient (angle between the x and y pixel axes).

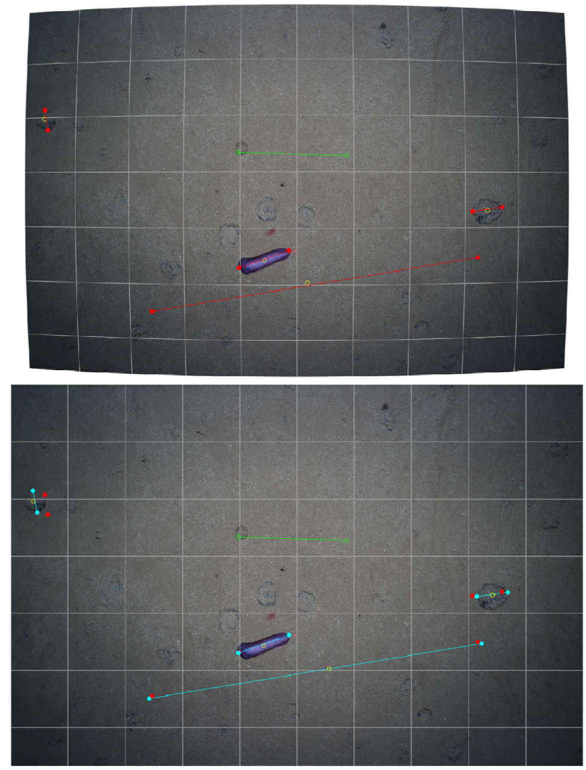

Fig. 4. Effect of lens distortion on the size measurements. Top: image artificially distorted with a symmetric barrel distortion ($k1 = 0.1$). Bottom: original undistorted image. The scale bar (green line) represents 50 cm. The red dots and lines are size measurements made on the distorted image (without lens correction). The blue dots and lines are the corresponding undistorted size measurements (with lens correction). Note that the error between the red and blue dots increases with distance from the image centre. In this example, the errors on the size measurements range from 1.6% (3 mm) for the sea cucumber near the centre to 10.4% (1.2 cm) for the nodule near the left edge of the image. Such errors can be easily corrected in *PAPARA(ZZ)I* by compensating for the lens distortion if the camera calibration parameters are known. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The camera intrinsic parameters and the lens radial and tangential distortion coefficients [63] must be provided in a text file called *camera_intrinsic.pap* and located in the same folder as the images. The structure of the *camera_intrinsic.pap* file is described in detail in the *PAPARA(ZZ)I* user manual, and an example file is provided with the *PAPARA(ZZ)I* test dataset.

2.2.7. Excluding images from the analysis and defining useable image areas

Due to strong absorption of light underwater, areas of deep-sea images are often too dark for analysis or impacted by strong lighting heterogeneities and vignetting. Reliable classification of the fauna in dark areas may be impaired. It is possible to exclude such areas from the analysis be setting the *useable area* of the image. By excluding parts of the images from the analysis, the selection of the useable area has an effect on the area calculation as well as the computed faunal abundances and densities. The useable area is set by the user for each image. If appropriate, the entire image can also be flagged as unusable.

2.2.8. Image enhancement

The brightness, contrast and gamma correction of images can be adjusted at any time during the analysis in order to enhance the image quality. This is done by using one of the three sliders located in the upper panel of the GUI (Fig. 5). This feature facilitates the analysis of images of poor quality.

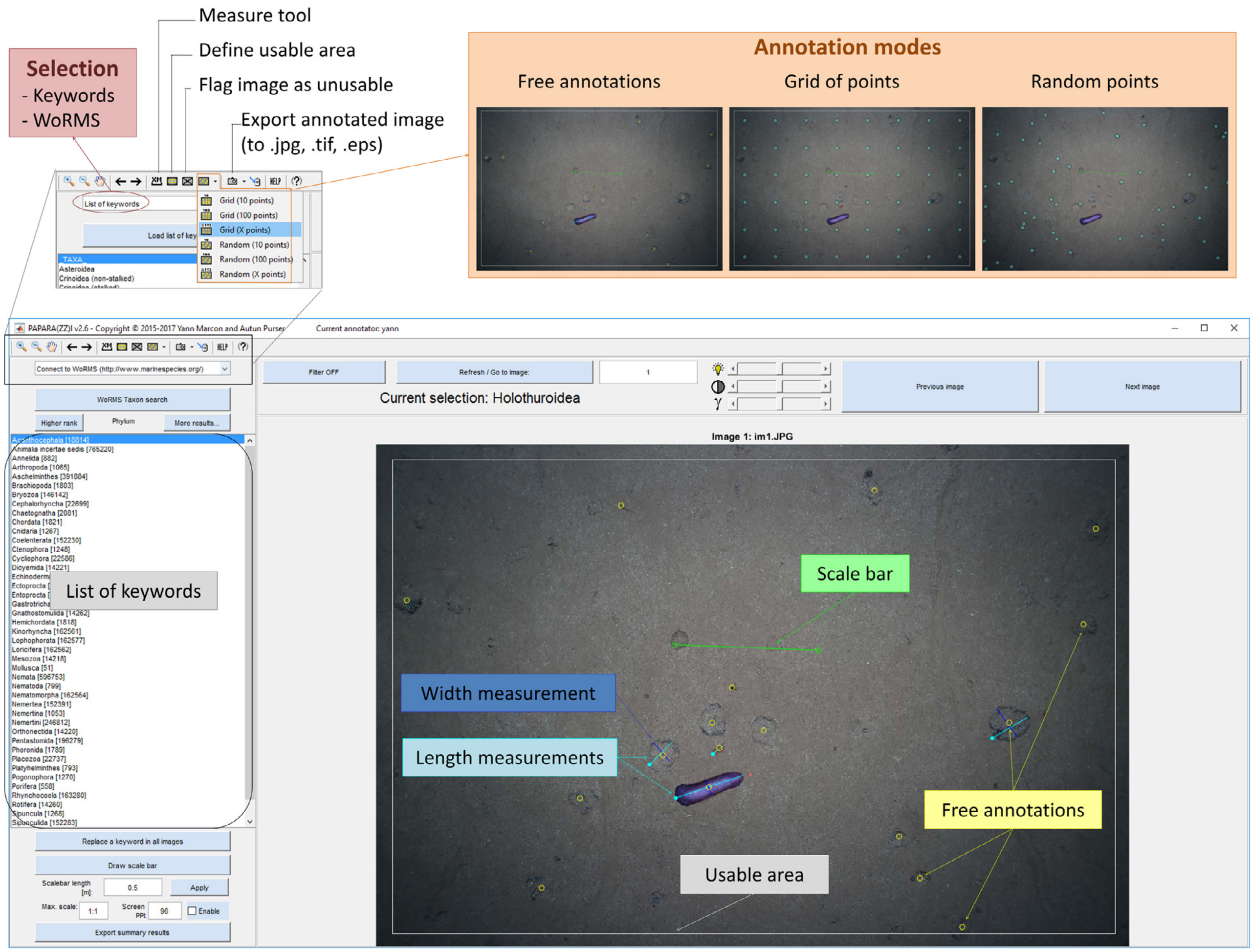

Fig. 5. View of the main annotation graphical interface of *PAPARA(ZZ)I*. Most of the functionalities can be called directly from this interface. The image shows free annotations (yellow circles), length/width measurements (light/dark blue lines), a scale bar (green line) and the boundary of the useable area (white rectangle). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2.9. Annotation filter

The annotation filter allows the user to display only certain categories of annotation when browsing through the images in *PAPARA(ZZ)I* (using the *previous image* and *next image* buttons). It also modifies the function of the *previous image* and *next image* buttons of the GUI. For instance, when the filter is *on*, the navigation buttons will automatically jump to the next image that contains one of the selected annotation categories. This function is useful to retrieve a particular annotation of interest (e.g. a rare one taxon throughout a large set of images), or to make modifications to certain types of annotations only (e.g. to correct for a species misclassified or to further subdivide a category).

2.2.10. Batch replacement of annotation attributes

Classifying marine fauna from underwater images is a difficult task and experts do not always agree on how to interpret their observations [45,59]. This fact implies that some observed taxa may sometimes be consistently mistaken for other taxa or be unidentifiable and labelled as *unknown taxa*. It may be a tedious and time-consuming task to correct all erroneous annotations once the culprit taxa have finally been identified. *PAPARA(ZZ)I* has the option to replace all selected keywords at once in all images of the open project instead of changing them all manually.

2.2.11. WoRMS, classification schemes and lists of keywords

PAPARA(ZZ)I is an official client of the web-services of the World Register of Marine Species [24]. The user can access and query the WoRMS register from the main interface of *PAPARA(ZZ)I*, and directly use the query results to annotate the images. Annotations made through the WoRMS services are automatically appended with the Aphia ID unique identifier [64], which ensures full compatibility of the annotations for future works. Indeed, the Aphia ID is a unique identifier for each taxon, which (1) enables the full use of the taxonomic hierarchical system benefitting studies that focus on certain taxonomic levels, and (2) facilitates the correlation of the annotations with results from other works.

Alternatively, *PAPARA(ZZ)I* can also be used with other existing classification schemes or with customised lists of keywords. Copies of the CATAMI Classification Scheme [22,23] and of the Coastal and Marine Ecological Classification Standard (CMECS) [20,21,65] specifically formatted for use with *PAPARA(ZZ)I* can be downloaded from the *PAPARA(ZZ)I* repository on GitHub (https://github.com/PAPARA-ZZ-I).

The CATAMI Classification Scheme is a standardised hierarchical classification system for annotating benthic biota and substratum types in underwater imagery. It takes into account the limitations of underwater images, especially for taxonomic identification, by suggesting a terminology based on coarse-level taxonomy and morphology [22,23]. As with the WoRMS-based annotations, the CATAMI-based annotations also include a unique identifier (Codes for Australian Aquatic Biota or CAAD) that facilitates comparison and integration into other synthesis works [22,66,67].

The *PAPARA(ZZ)I* version of the CMECS classification scheme contains only the “biotic” and “substrate” components of the classification standard [21]. It omits the "water column" and "geoform" components, which are not suited for megafauna analyses based on point annotations, and describe aspects of the marine environment that are beyond the scope of the program. The CMECS catalogue also uses its own unique identifiers, which are appended to the annotations made with *PAPARA(ZZ)I*.

Customised lists of keywords can also be created by the users with a simple text editor. A list of keywords is a simple text file (*.txt) with one keyword per line. Keywords can be organised into groups separated by empty lines. Group titles (i.e. non-selectable keywords) can be defined by inserting a "_" character before and after the keyword (example: *_Group Title_*). There is no limit to the number of keywords or groups. An example "list of keywords" file is provided with the *PAPARA(ZZ)I* test dataset.

2.2.12. Export annotated image

Each annotated image (e.g. the image including points, scale bar and limits of useable area) can be exported and saved in different file formats (*.jpeg, *.tiff, *.eps). This is done directly from the main GUI and the images are automatically stored in a subdirectory ('*username*_exported_images\\') of the project folder. This is convenient to show annotated images in presentations or for use in publications. The *.eps formats (EPS level 1 and 2) are vector-based and give the possibility to edit annotations in graphic design programs (e.g. Illustrator, CorelDraw, Inkscape, etc.) for figure preparation.

2.2.13. Data export

PAPARA(ZZ)I writes all raw data into text files but does not perform any calculation or measurement during the image analysis process. Raw data include the pixel coordinates of every point and size measurements of measured features as well as the associated annotation text. All additional calculations are done internally during the data export. Data export can be done at any time during analysis. For instance, the user can check preliminary results after just a few images have been analysed. Export is done by a single button click and may take a few minutes to complete, depending on the number of images and annotations made, or if the lens correction functionality was applied or not.

Exported data are stored in several tab-delimited text files and are structured for display in a spreadsheet program such as Excel. The tab-delimited text file format was chosen because (1) it can be read natively by most text editors, spreadsheet and statistical analysis programs, (2) it can be parsed easily for data manipulation (MATLAB, R, Python, C, etc.), and (3) it does not have the disadvantages of comma-delimited (.csv) files (for instance if annotation texts contain commas). The output files include ecological indices (abundances per image, abundances per taxon, total per-taxon abundances, and total per-taxon densities) as well as a selection of statistical values based on the size measurements (sorted by taxa). These data allow the user to easily construct abundance/sample matrices to conduct further statistical analyses [68]. Additionally, the per-taxon size distribution data series as well as the direction angles of the size measurements are also exported. All values are exported for both the whole area and the useable area of the images.

3. Illustrative examples

In this section we describe the main user graphical interface of *PAPARA(ZZ)I* and illustrate some of the functionalities that were described in Section 2.2. A user manual is provided with the program that describes how to use each functionality in full detail.

Upon starting, *PAPARA(ZZ)I* prompts the user for a username, for the path to the project folder (i.e. the folder that contains the images to annotate), and to decide whether to randomise the images before the main GUI appears. All the program functionalities (described in Section 2.2) – except the image order randomisation – can be called directly from the main interface (Fig. 5). Before starting the annotation process, the user must either select a so-called *list of keywords* (which can be a classification scheme) or connect to the WoRMS database. At this point, the annotation interface becomes enabled. If the WoRMS web-services are activated, it is possible to search for a particular taxon through the "WoRMS Taxon search" menu. It gives the possibility to search and display information about a particular taxon. Taxa can be searched by scientific name, common name, Aphia ID or a wide range of identifiers commonly used in taxonomy (Fig. 6).

During the annotation process, the user can browse through the images of the project with the navigation buttons, go directly to a desired image, filter the annotations, or flag an image as unusable (Fig. 7). Buttons to switch to other annotation modes (grid of points or random points) are available in the main toolbar at the top of the GUI (Fig. 5). The naming convention of the annotation files allows the program to keep track of all annotations made with the different annotation modes. Therefore, users can freely switch back and forth between annotation modes without losing any annotations. This can be useful when trying to determine, for instance, whether a coarse 10 point grid, or more detailed but labour intensive 100 point grid, gives the level of analysis resolution required for a particular research task. Some tasks may require both a coarse analysis of the seafloor habitat type (such as 'muddy seafloor', 'rock' etc.) and the logging of all individuals of a particular fauna type [69]. Such analysis can be done by using a point grid to analyse the habitat type and the free annotation mode to log the organisms. *PAPARA(ZZ)I* allows both annotation modes to be carried out within the same analysis task.

Results can be exported to summary files anytime by simply selecting the desired annotation mode (i.e. free points, point grid, random points) and then the "Export summary results" button. This function exports only the results of the selected annotation mode. Results of other annotation modes can be exported by repeating this procedure with another mode. All results are exported to a subdirectory of the project folder called 'data-export'. Within this subdirectory, the results are sorted in further subdirectories named after the date/time of data export and the annotator's name, as well as the annotation mode in order to prevent overwriting of any previously exported result. The resulting summary files are formatted for optimal display in Excel, but output files can also be read and manipulated with other data analysis programs such as R or MATLAB (Fig. 1). The summary files show the faunal counts per taxon and per image, the density of each taxon over the entire project, and the results of the size measurements (Fig. 7).

There is no button to enable or disable the correction for lens distortion. The correction is applied automatically if the file that contains the camera intrinsic and lens distortion parameters is present in the project folder. More information about this file can be found in the user manual and an example file is supplied with the *PAPARA(ZZ)I* test dataset. It is possible to look at the differences between corrected and uncorrected results by exporting the results twice: once with and once without the parameter file placed within the project folder. In the example shown in Fig. 4, the errors on the size measurements ranged from 1.6% near the centre to 10.4% near the edge of the image, which in world units represents errors of 0.3 to 1.2 cm. These values are significant considering that the photo was taken at an altitude of 1.5 m above seafloor. Seabed photo surveys are often conducted at higher altitudes (up to 10 m) [70–72], further increasing the magnitude of such errors.

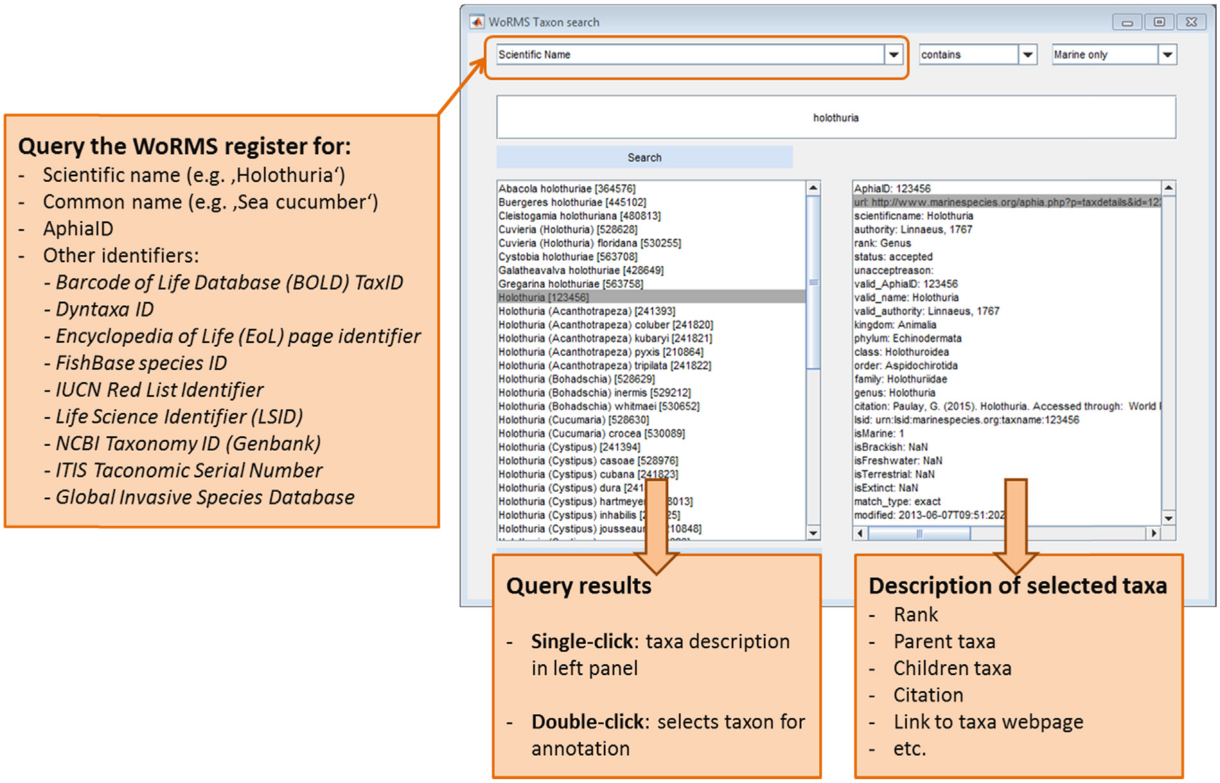

Fig. 6. The WoRMS Taxon search interface of *PAPARA(ZZ)I*. Taxa names can be queried in various ways and information about each taxon can be displayed easily. Once the desired taxon has been found, it is possible to directly use it for annotating the images.

4. Impact

The development of *PAPARA(ZZ)I* started in September 2015 during the cruise SO242/2 to the DISCOL experimental area [26,73,74]. The initial drive for developing this program was to respond to an immediate need for an image annotation routine that could be used by several people to work on the same image datasets without Internet access. Feedbacks from the first users motivated the further development of *PAPARA(ZZ)I* into an end-user product that provides marine ecologists with an easy-to-use and reliable tool for annotating large sets of images and summarising the results in an automatic way. Presently, *PAPARA(ZZ)I* has been used or tested internationally by scientists in Germany (Alfred Wegener Institute, MARUM) [74], Belgium (Ghent University), United Kingdom (National Oceanography Centre) and USA (Florida State University) for analysing both deep-sea and shallow coastal benthic communities. It is also used to teach students of the ROBEX-funded web course on Image Analysis and Spatial Statistics (IASS) of the Jacobs University (Bremen, Germany) [75,76]. Input from these users helped improve the available functionalities of *PAPARA(ZZ)I* based on real world necessities. Several functionalities, such as the facility to query the WoRMS register, to allow selection of a useable area of an image and the compatibility with different operating systems (Windows, Mac OS X, Linux) were implemented in *PAPARA(ZZ)I* following user comments and wishes.

4.1. A lightweight and comprehensive solution for benthic photo studies

PAPARA(ZZ)I is a lightweight and easy to set up solution, which offers all functionalities that are required for benthic ecological studies based on point annotations of photographs. It provides an efficient means to annotate a large number of images for marine ecological studies, as well as for other types of studies based on point annotations of features. It is fully functional on Windows, Linux and Mac OS X and all output data generated by the program are fully compatible with all operating systems.

PAPARA(ZZ)I changes the daily practice of its users in that they can (1) work on any operating system and from anywhere even without Internet connection, (2) conduct various types of point annotation techniques with the same interface (consistency of data- and work-flow between projects), (3) start annotating images immediately after acquisition without the need of uploading them onto a web-based tool or server, (4) split large image projects between several users and easily re-combine results in order to produce summary files (calculation of abundances, densities and size statistics, etc.) for the entire project at once, (5) reduce analysis biases by randomising the image order and limiting the maximum display scale for all users analysing the same dataset, (6) reduce error in summary calculations, (7) directly reuse raw annotation and result files in other programs for further analyses of the data (statistics, ecological indices, habitat maps, etc.) and, (8) share annotation files between users and provide transparent and repeatable results to the scientific community.

Furthermore, *PAPARA(ZZ)I* gives scientists full control and freedom to work on their datasets without the need to rely on a third party or the need for a larger server infrastructure. Not only does *PAPARA(ZZ)I* allow users to analyse images without any of those constraints, but it also gives them the possibility to start the annotation process even during the image acquisition survey. Indeed, images can be added to a *PAPARA(ZZ)I* project at any time, simply by copying them into the project folder and by restarting *PAPARA(ZZ)I*. This is hugely beneficial in the research cruise environment, allowing the commencement of image data analysis whilst still at sea, not requiring a wait until the end of the survey and fast Internet access before being able to start the analysis – a considerable saving in time and advantageous for fine-tuning ongoing data collection efforts.

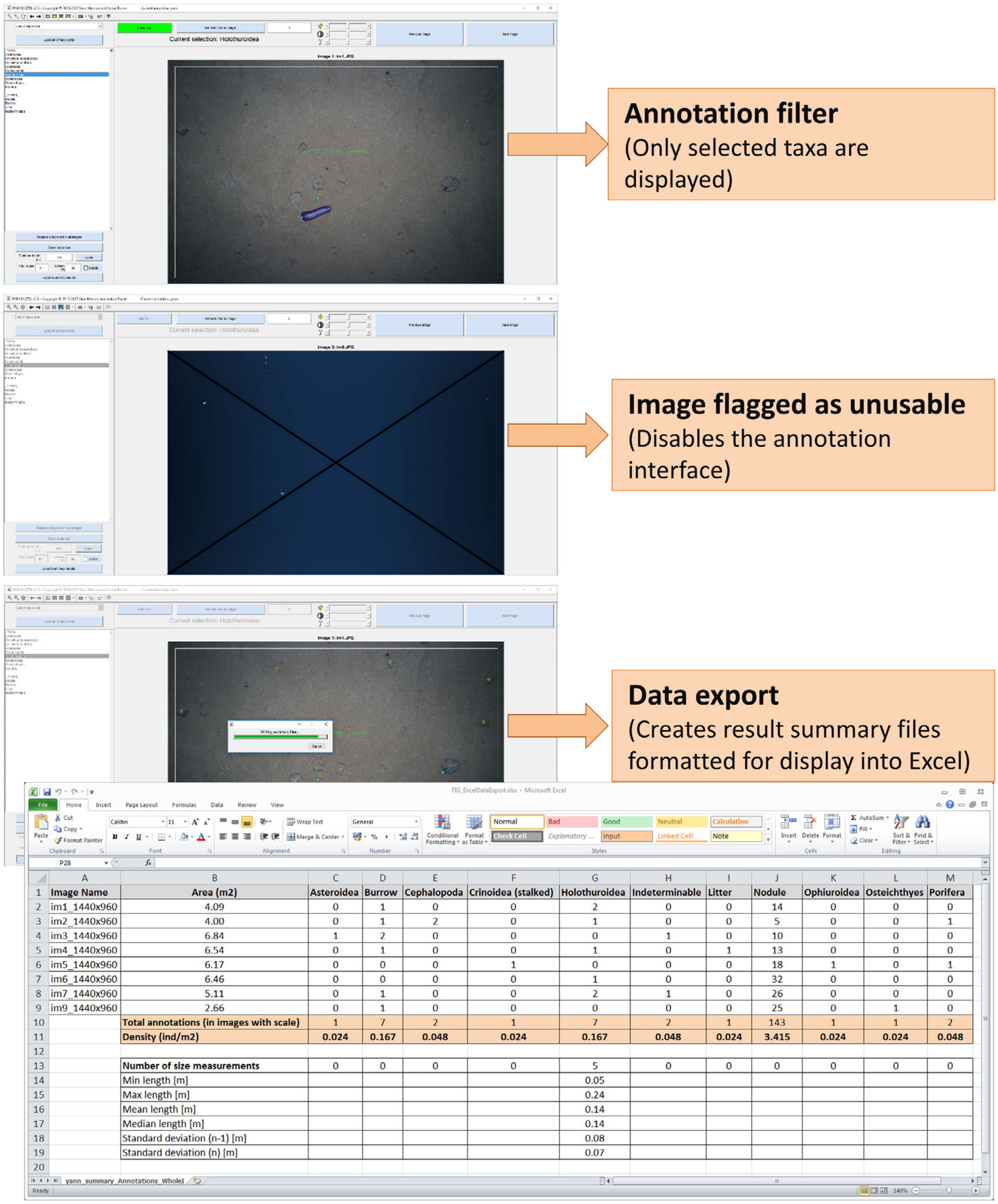

Image Name	Area (m2)	Asteroidea	Burrow	Cephalopoda	Crinoidea (stalked)	Holothuroidea	Indeterminable	Litter	Nodule	Ophiuroidea	Osteichthyes	Porifera
im1_1440x960	4.09	0	1	0	0	2	0	0	14	0	0	0
im2_1440x960	4.00	0	1	2	0	1	0	0	5	0	0	1
im3_1440x960	6.84	1	2	0	0	0	1	0	10	0	0	0
im4_1440x960	6.54	0	1	0	0	1	0	1	13	0	0	0
im5_1440x960	6.17	0	0	0	1	0	0	0	18	1	0	1
im6_1440x960	6.46	0	0	0	0	1	0	0	32	0	0	0
im7_1440x960	5.11	0	1	0	0	2	1	0	26	0	0	0
im9_1440x960	2.66	0	1	0	0	0	0	0	25	0	1	0
	Total annotations (in images with scale)	1	7	2	1	7	2	1	143	1	1	2
	Density (ind/m2)	**0.024**	**0.167**	**0.048**	**0.024**	**0.167**	**0.048**	**0.024**	**3.415**	**0.024**	**0.024**	**0.048**
	Number of size measurements	0	0	0	0	5	0	0	0	0	0	0
	Min length [m]					0.05						
	Max length [m]					0.24						
	Mean length [m]					0.14						
	Median length [m]					0.14						
	Standard deviation (n-1) [m]					0.08						
	Standard deviation (n) [m]					0.07						

Fig. 7. Additional functionalities allow the user to filter annotations when browsing through the project images, to remove unusable images from the analysis (images and annotations are not deleted, they are simply ignored by *PAPARA(ZZ)I* during the data export), and to export the results into table-structured text files for display in Excel.

4.2. Ease of use, consistency, transparency and versatility

PAPARA(ZZ)I does not claim to allow users to pursue new research questions but rather to streamline and speed up the workflow of ecological studies based on point annotations. It is easy to use thanks to a simple graphical user interface that contains only a few buttons and allocates as much screen space as possible for image display. All essential features for studies on marine megafauna diversity and abundance are present and they require only minor parameter input from the user. *PAPARA(ZZ)I* project files have a simple and clear data structure, which ensures transparency of the results. A key feature of the program is the simplicity and readability of its output files (raw annotations, useable area, scale bar, and result summaries). For instance, the raw annotation files contain the following basic information for each annotation: pixel position of the annotated feature on the image frame, annotation text, and pixel coordinates of both ends of the corresponding size measurement (if any). If needed, the data can be easily manipulated by the user with other programs or scripts, for instance for assessing localised abundance patterns (clustering, patchiness, spacing, etc.). *PAPARA(ZZ)I* is also reliable in that the raw data are saved at time of entry, with all subsequent calculations done during the data export. This ensures that the data export is uniform for every project, and that users can easily make corrections to the projects and re-export the results at any time with a single mouse click. The resulting summary files (Fig. 2) are produced automatically, which limits the risk of errors [50] by ensuring consistency of the data calculations between projects.

Finally, despite its simplicity of use, *PAPARA(ZZ)I* is not a black box program. The source code is open and users with programming experience can check and quality control calculations.

PAPARA(ZZ)I can be used with several existing classification schemes [20–23] as well as with user-defined customised lists of keywords. Not only does this make the program more versatile, but it also facilitates the annotation process. Customised lists are useful to speed up the annotation process by keeping the lists as short as possible, with only the keywords and taxa names that are of interest for the open project displayed. If it becomes apparent that a keyword is missing from the list in use, it can be added to the text file and the list can be reloaded into *PAPARA(ZZ)I* at any time during analysis. In most cases, customised lists are based on accepted taxa names from the WoRMS register or other classification schemes. However, keyword lists can also contain project-specific keywords that are not included in existing classification schemes. This facility broadens the range of applications of *PAPARA(ZZ)I* to studies not related to marine ecology and marine geosciences. For instance, keywords can include words that do not describe fauna or substrates (e.g. anthropogenic features, litter, etc.).

Customised lists come in handy in studies that compare faunal abundances and benthic community composition with the results of previous published work [50], allowing users to use the same faunal categories in order to make the analysis faster and comparison of results more direct [73]. These possibilities facilitate collaboration by training other scientists and students to carry out some of the analyses via *PAPARA(ZZ)I* at different institutes, and integrating their output files and results. This is especially useful for extremely large image datasets.

5. Conclusions

The favourable interest in *PAPARA(ZZ)I* shows that it responds to a need within the scientific community. This interest results from the fact that *PAPARA(ZZ)I* is easy and straightforward to use, yet powerful, considering the range of functionalities offered. Scientists can keep full control of their datasets as well as having the freedom to work from anywhere and on any computer, without the need to rely on a third party or the availability of a large server infrastructure.

Additional features may be implemented in future releases of *PAPARA(ZZ)I* to broaden the range of applications for use in the field of marine ecology and to further streamline the workflow. Functionalities that are currently considered include: patch annotations (to annotate extended spatial features, e.g. coral reef colonies or large sponges), calculation of area percentages based on point grids and on patch annotations to estimate cover area (e.g. percentage of gravel, sediments, coral cover etc.), the import of underwater positioning data to allow the exported *PAPARA(ZZ)I* annotation data to be georeferenced directly (in order to reduce post-processing steps shown in Fig. 1), the possibility to scale the images based on either the scale bar or from altitude data (for cases where images do not contain laser points or other features used for scaling but the flight height of the vehicle above seafloor is known). Further calculations that will be considered in the data export include the following: automatic production of inter-project abundance/sample matrices (for inter-survey comparisons) and calculations of common ecological indices (richness, Shannon diversity index, Bray–Curtis index, species similarity, patchiness etc.) to reduce post-processing steps for the users. However, the further development of *PAPARA(ZZ)I* should not be restricted to the research aims of its initial creators and, by publishing *PAPARA(ZZ)I* under an open source licence, we hope to encourage contributions from other scientists and developers. Finally, we wish to make *PAPARA(ZZ)I* fully compatible with GNU Octave in order to render possible the use of the source code by users who do not have access to MATLAB.

Acknowledgements

The authors would like to thank the chief scientist, captain and crew of RV Sonne, as well as Harald Biebow for their support during cruise SO242/2. Thanks also to Lidia Lins (Ghent University), Melanie Bergmann (Alfred Wegener Institute), Nicole Morgan and Caroline Johansen (Florida State University) for their feedback and suggestions. Thanks to the anonymous reviewers for their useful and constructive comments. The SO242/2 cruise was funded by the German Ministry of Education and Science BMBF (grant number 03F0707A-G) through the project Mining Impact of the Joint Programming Initiative Healthy and Productive Seas and Oceans (JPIO). This work was supported by the Helmholtz Alliance "Robotic Exploration of Extreme Environments – ROBEX" (contract no. HA-304) and the European Union Seventh Framework Programme (FP7/2007-2013) under the MIDAS project, grant agreement no. 603418.

References

[1] Kato Y, et al. Deep-sea mud in the pacific ocean as a potential resource for rare-earth elements. Nat Geosci 2011;4(8):535–9. http://dx.doi.org/10.1038/ngeo1185.

[2] Copley JTP. Just how little do we know about the ocean floor? The Conversation, 09-Oct-2014.

[3] Glover AG, et al. Temporal change in deep-sea benthic ecosystems: a review of the evidence from recent time-series studies. Adv Mar Biol 2010;58:1–95. http://dx.doi.org/10.1016/B978-0-12-381015-1.00001-0.

[4] Roberts CM. Deep impact: the rising toll of fishing in the deep sea. Trends Ecol Evol 2002;17(5):242–5. http://dx.doi.org/10.1016/S0169-5347(02)02492-8.

[5] Barbier EB, et al. Ecology: Protect the deep sea. Nature 2014;505(7484):475–7. http://dx.doi.org/10.1038/505475a.

[6] Lodge M, Johnson D, Le Gurun G, Wengler M, Weaver P, Gunn V. Seabed mining: International Seabed Authority environmental management plan for the Clarion–Clipperton Zone. A partnership approach. Mar Policy 2014;49: 66–72. http://dx.doi.org/10.1016/j.marpol.2014.04.006.

[7] Mengerink KJ, et al. A call for deep-ocean stewardship. Science 2014; 344(6185):696–8. http://dx.doi.org/10.1126/science.1251458.

[8] Moskvitch K. Health check for deep-sea mining. Nature 2014;512(7513): 122–3. http://dx.doi.org/10.1038/512122a.

[9] Wedding LM, et al. Managing mining of the deep seabed. Science 2015; 349(6244):144–5. http://dx.doi.org/10.1126/science.aac6647.

[10] McCauley DJ, Pinsky ML, Palumbi SR, Estes JA, Joyce FH, Warner RR. Marine defaunation: Animal loss in the global ocean. Science 2015;347(6219): 1255641. http://dx.doi.org/10.1126/science.1255641.

[11] Gage JD, Tyler PA. Deep-sea biology: a natural history of organisms at the deep-sea floor. Cambridge University Press; 1991.

[12] Schlining BM, Stout NJ. MBARI's video annotation and reference system. In: OCEANS 2006, Boston, MA, 2006, p. 1–5. http://dx.doi.org/10.1109/OCEANS.2006.306879.

[13] 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. Comput Geosci 2006;32(9):1259–69. http://dx.doi.org/10.1016/j.cageo.2005.11.009.

[14] 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 2009 - EUROPE, 2009, p. 1–7. http://dx.doi.org/10.1109/OCEANSE.2009.5278332.

[15] CoralNet. [Online]. Available: http://coralnet.ucsd.edu/. [Accessed 16 December 2015].

[16] Ayroza C, Hustache J, Fagundes L, Sumi M, Ferrari R. Finding the right tool for benthic image analysis. Reef Encount Int Soc Reef Stud 2015;30:34.

[17] Gomes-Pereira JN, et al. Current and future trends in marine image annotation software. Prog Oceanogr 2016;149:106–20. http://dx.doi.org/10.1016/j.pocean.2016.07.005.

[18] Ferrini VL, Singh H. FISH_ROCK: A tool for identifying and counting benthic organisms in bottom photographs, Woods Hole Oceanographic Institution, Woods Hole, Massachussetts, Technical report WHOI-2006-01, Jan. 2006.

[19] Abramoff MD, Magalhães PJ, Ram SJ. Image processing with imagej. Biophotonics Int 2004;11(7):36–42.

[20] Madden CJ, et al. Coastal and marine ecological classification standard. Natl Ocean Atmospheric Adm NatureServe 2009;77.

[21] Coastal and Marine Ecological Classification Standard, Marine and Coastal Spatial Data Subcommittee, Federal Geographic Data Committee, FGDC-STD-018-2012, Jun. 2012.

[22] Althaus F, Hill NA, Edwards L, Ferrari R. CATAMI Classification Scheme for scoring marine biota and substrata in underwater imagery - A pictorial guide to the Collaborative and Annotation Tools for Analysis of Marine Imagery and Video (CATAMI) classification scheme; 2013. [Online]. Available: http://catami.org/classification. [Accessed 04 December 2015].
[23] Althaus F, et al. A standardised vocabulary for identifying benthic biota and substrata from underwater imagery: the CATAMI classification scheme. PLOS ONE 2015;10(10):e0141039. http://dx.doi.org/10.1371/journal.pone.0141039.
[24] Boxshall GA. et al. World Register of Marine Species (WoRMS). WoRMS Editorial Board, 04-Dec-2015.
[25] GNU General Public License version 3, GNU; 2007. [Online]. Available: http://www.gnu.org/copyleft/gpl.html. [Accessed 25 December 2015].
[26] Boetius A. Short cruise report RV SONNE – SO242/2 (28.08.2015 - 01.10.2015). In: Max Planck institute for marine microbiology, Bremen, Cruise Report SO242/2, 2015.
[27] Diepenbroek M, et al. PANGAEA—an information system for environmental sciences. Comput Geosci 2002;28(10):1201–10. http://dx.doi.org/10.1016/S0098-3004(02)00039-0.
[28] Grobe H, Diepenbroek M, Dittert N, Reinke M, Sieger R. Archiving and distributing earth-science data with the PANGAEA information system. In: Fütterer PDDK, Damaske DD, Kleinschmidt PDG, Miller PDDhcH, Tessensohn DF, editors. Antarctica. Berlin Heidelberg: Springer; 2006. p. 403–6. http://dx.doi.org/10.1007/3-540-32934-X_51.
[29] British Oceanographic Data Centre (BODC), Liverpool, Merseyside, UK, British Oceanographic Data Centre. [Online]. Available: http://www.bodc.ac.uk/. [Accessed 08 December 2015].
[30] Long Term Ecological Research Network (LTER). LTER network data portal. [Online]. Available: https://portal.lternet.edu/nis/home.jsp [Accessed 8 December 2015].
[31] Marine Geoscience Data System (MGDS). Marine geoscience data system. [Online]. Available: http://www.marine-geo.org. [Accessed 8 December 2015].
[32] National Centers for Environmental Information (NCEI). National centers for environmental information (NCEI). [Online]. Available: https://www.ncei.noaa.gov/. [Accessed 8 December 2015].
[33] PANGAEA: Data Publisher for Earth & Environmental Science, *PANGAEA*. [Online]. Available: http://www.pangaea.de/. [Accessed 08 December 2015].
[34] Pielou EC. Population and community ecology: Principles and methods. Gordon & Breach. New York: CRC Press; 1974.
[35] Andrew NL, Mapstone BD. Sampling and the description of spatial pattern in marine ecology. Oceanogr Mar Biol Annu Rev 1987;25:39–90.
[36] Dethier MN, Graham ES, Cohen S, Tear LM. Visual versus random-point percent cover estimations: 'objective' is not always better. Mar Ecol Prog Ser 1993; 96(9):93–100.
[37] Carleton JH, Done TJ. Quantitative video sampling of coral reef benthos: large-scale application. Coral Reefs 1995;14(1):35–46. http://dx.doi.org/10.1007/BF00304070.
[38] Brown EK, et al. Development of benthic sampling methods for the coral reef assessment and monitoring program (CRAMP) in hawai'i. Pac Sci 2004;58(2): 145–58.
[39] Roelfsema CM, Phinn SR, Joyce KE. Evaluating benthic survey techniques for validating maps of coral reefs derived from remotely sensed images. In: Proc 10th int coral reef symp, vol. 1. 2006. p. 1771–80.
[40] Leujak W, Ormond RFG. Comparative accuracy and efficiency of six coral community survey methods. J Exp Mar Biol Ecol 2007;351(1–2):168–87. http://dx.doi.org/10.1016/j.jembe.2007.06.028.
[41] Guinan J, Grehan AJ, Dolan MFJ, Brown C. Quantifying relationships between video observations of cold-water coral cover and seafloor features in rockall trough, west of ireland. Mar Ecol Prog Ser 2009;375:125–38. http://dx.doi.org/10.3354/meps07739.
[42] Purser A, Bergmann M, Lundlv T, Ontrup J, Nattkemper TW. Use of machine-learning algorithms for the automated detection of cold-water coral habitats: a pilot study. Mar Ecol Prog Ser 2009;397:241–51. http://dx.doi.org/10.3354/meps08154.
[43] Smale DA, Kendrick GA, Waddington KI, Van Niel KP, Meeuwig JJ, Harvey ES. Benthic assemblage composition on subtidal reefs along a latitudinal gradient in western australia. Estuar Coast Shelf Sci 2010;86(1):83–92. http://dx.doi.org/10.1016/j.ecss.2009.10.016.
[44] Bewley M, et al. Australian sea-floor survey data, with images and expert annotations. Sci Data 2015;2:150057. http://dx.doi.org/10.1038/sdata.2015.57.
[45] Schoening T, 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. http://dx.doi.org/10.1371/journal.pone.0038179.
[46] Foell EJ, Pawson DL. Photographs of invertebrate megafauna from abyssal depths of the north-eastern equatorial pacific ocean, Jun. 1986.
[47] Bluhm H. Monitoring megabenthic communities in abyssal manganese nodule sites of the East Pacific Ocean in association with commercial deep-sea mining. Aquat Conserv Mar Freshw Ecosyst 1994;4(3):187–201. http://dx.doi.org/10.1002/aqc.3270040302.
[48] Bluhm H, Schriever G, Thiel H. Megabenthic recolonization in an experimentally disturbed abyssal manganese nodule area. Mar Georesources Geotechnol 1995;13(4):393–416. http://dx.doi.org/10.1080/10641199509388295.
[49] Radziejewska T, Stoyanova V. Abyssal epibenthic megafauna of the clarion-clipperton area (NE Pacific): Changes in time and space versus anthropogenic environmental disturbance. Oceanol Stud 2000;29(2):83–101.
[50] Bluhm H. Re-establishment of an abyssal megabenthic community after experimental physical disturbance of the seafloor. Deep Sea Res Part II Top Stud Oceanogr 2001;48(17–18):3841–68. http://dx.doi.org/10.1016/S0967-0645(01)00070-4.
[51] De Leo FC, Smith CR, Rowden AA, Bowden DA, Clark MR. Submarine canyons: hotspots of benthic biomass and productivity in the deep sea. Proc R Soc Lond B Biol Sci 2010.
[52] Bergmann M, Soltwedel T, Klages M. The interannual variability of megafaunal assemblages in the Arctic deep sea: Preliminary results from the HAUSGARTEN observatory (79° N). Deep Sea Res Part Oceanogr Res Pap 2011;58(6):711–23. http://dx.doi.org/10.1016/j.dsr.2011.03.007.
[53] Bergmann M, Klages M. Increase of litter at the arctic deep-sea observatory HAUSGARTEN. Mar Pollut Bull 2012;64(12):2734–41. http://dx.doi.org/10.1016/j.marpolbul.2012.09.018.
[54] Duffy GA, Lundsten L, Kuhnz LA, Paull CK. A comparison of megafaunal communities in five submarine canyons off southern california, USA. Deep Sea Res Part II Top Stud Oceanogr 2014;104:259–66. http://dx.doi.org/10.1016/j.dsr2.2013.06.002.
[55] Durden JM, Bett BJ, Jones DOB, Huvenne VAI, Ruhl HA. Abyssal hills – hidden source of increased habitat heterogeneity, benthic megafaunal biomass and diversity in the deep sea. Prog Oceanogr A 2015;137:209–18. http://dx.doi.org/10.1016/j.pocean.2015.06.006.
[56] Speight M, Henderson PA. Marine ecology: Concepts and applications. Wiley-Blackwell; 2010.
[57] Boetius A. et al. Short cruise report - MERIAN MSM16/2 - 'LOOME'- long term observatory of mud volcano eruptions, Alfred Wegener Institut für Polar und Meeresforschung, Bremerhaven, Cruise Report, 2010.
[58] Howell KL, Bullimore RD, Foster NL. Quality assurance in the identification of deep-sea taxa from video and image analysis: response to henry and roberts. ICES J Mar Sci J Cons 2014;71(4):899–906. http://dx.doi.org/10.1093/icesjms/fsu052.
[59] Beijbom O, et al. Towards automated annotation of benthic survey images: Variability of human experts and operational modes of automation. PLoS ONE 2015;10(7):e0130312. http://dx.doi.org/10.1371/journal.pone.0130312.
[60] Brown DC. Decentering distortion of lenses. Photom Eng 1966;32(3):444–62.
[61] Hartley R, Zisserman A. Multiple view geometry in computer vision. Cambridge University Press; 2003.
[62] A. Zhang. A flexible new technique for camera calibration. IEEE Trans Pattern Anal Mach Intell 2000;22(11):1330–4.
[63] Heikkila J, Silven O. A four-step camera calibration procedure with implicit image correction. In: Proceedings of the 1997 IEEE computer society conference on computer vision and pattern recognition, San Juan, 1997, p. 1106–12.
[64] Vandepitte L, et al. How aphia—the platform behind several online and taxonomically oriented databases—can serve both the taxonomic community and the field of biodiversity informatics. J Mar Sci Eng 2015;3(4):1448–73. http://dx.doi.org/10.3390/jmse3041448.
[65] CMECS: Coastal and Marine Ecological Classification Standard, 2012. [Online]. Available: http://www.cmecscatalog.org/. [Accessed 12 December 2015].
[66] Yearsley GK, Last PR, Morris GB. Codes for Australian aquatic biota (CAAB): an upgraded and expanded species coding system for Australian fisheries databases, CSIRO Marine Laboratories, 224, 1997.
[67] Rees AJJ, Yearsley GK, Gowlett-Holmes K, Pogonoski J. Codes for Australian Aquatic Biota (on-line version), CSIRO Marine Research, World Wide Web electronic publication, 1999. [Online]. Available: http://www.cmar.csiro.au/caab/.
[68] Clarke KR, Warwick RM. Change in Marine communities: An approach to statistical analysis and interpretation, vol. 2. Plymouth, UK: PRIMER-E; 2001.
[69] Purser A, et al. Microhabitat and shrimp abundance within a Norwegian cold-water coral ecosystem. Biogeosciences 2013;10(9):5779–91. http://dx.doi.org/10.5194/bg-10-5779-2013.
[70] Bluhm H. Re-establishment of an abyssal megabenthic community after experimental physical disturbance of the seafloor. Deep Sea Res Part II Top Stud Oceanogr 2001;48(17–18):3841–68. http://dx.doi.org/10.1016/S0967-0645(01)00070-4.
[71] Escartín J, et al. Globally aligned photomosaic of the Lucky Strike hydrothermal vent field (Mid-Atlantic Ridge, 37° 18.5'N): Release of georeferenced data, mosaic construction, and viewing software. Geochem Geophys Geosystems 2008;9(12):17. http://dx.doi.org/10.1029/2008GC002204.
[72] Marcon Y, Ondréas H, Sahling H, Bohrmann G, Olu K. Fluid flow regimes and growth of a giant pockmark. Geology 2014;42(1):63–6. http://dx.doi.org/10.1130/G34801.1.
[73] Purser A, Marcon Y, Boetius A. Megabenthic community structure within and surrounding the DISCOL experimental area 26 years after simulated manganese nodule mining disturbance, presented at the 2016 Ocean Science Meeting, New Orleans, Louisiana, USA, 21-Feb-2016.
[74] Purser A, et al. Association of deep-sea incirrate octopods with manganese crusts and nodule fields in the pacific ocean. Curr Biol 2016;26(24):R1268–9. http://dx.doi.org/10.1016/j.cub.2016.10.052.
[75] Williams R, Purser A, Lund S. Building inter-disciplinary competence in images analysis and spatial statistics though collaborative online learning and remote field trips. INTED2014 Proc 2014;1029–37.
[76] Purser A, Bamberg M, Thomsen L. Robotic exploration of extreme environments: Image analysis and spatial statistics (IASS). [Online]. Available: http://imageanalysis.weebly.com/. [Accessed 12 January 2016].