photoQuad: A dedicated seafloor image processing software, and a comparative error analysis of four photoquadrat methods

V. Trygonis a,b, M. Sini a,b

a Fisheries Management & Fisheries Acoustics Laboratory, Department of Marine Sciences, University of the Aegean, Mytilene, 81100, Greece
b Marine Biodiversity & Coastal Management Laboratory, Department of Marine Sciences, University of the Aegean, Mytilene, 81100, Greece

ARTICLE INFO

Article history:
Received 5 January 2012
Received in revised form 6 April 2012
Accepted 30 April 2012
Available online 8 June 2012

Keywords:
Coralligenous
Benthic communities
Coverage measurement
Image processing software
Photoquadrat sampling
Simulation

ABSTRACT

Photographic quadrat sampling is commonly used for the study of sessile benthic communities. However, photoquadrat data handling is fragmented into several processing methods, and there is a scarcity of dedicated software tools that integrate all major analysis options. photoQuad is a new software system for advanced image processing of photographic samples, dedicated to ecological applications. The software integrates a series of methods for the extraction of species area, percentage coverage, or presence/absence information, including random point counts (RP), grid cell counts (CL), freehand regions (FH), and image segmentation-based regions (SG). These are simultaneously functional in a layer-based environment, further supported by a variety of tools for image enhancement, image calibration, automatic quadrat boundary detection, and management of user-specific species libraries. The paper documents the main features of photoQuad, and demonstrates its performance through the simultaneous application of the RP, CL, FH, and SG methods on identical datasets, and the comparison of errors in species area and coverage measurements. The simulated data used for reference are disk-shaped patches, whose area and density statistics are equivalent to three benthic species characteristic of Mediterranean coralligenous communities. The analysis indicated that measurement methods differed in area and coverage bias, as well as in their sensitivity to species size. Large patches were accurately measured by all methods in terms of mean scaled error, but CL and RP provided high error variance, and their performance deteriorated with decreasing patch size; the region-based SG and FH methods provided the lowest errors and were both robust to patch size. The image and quadrat calibration process showed no statistically significant effect on the outputs, although further analysis is needed to validate this result. Overall, photoQuad constitutes a powerful software for elaborate analysis of photoquadrat images, facilitating fast and comparable evaluation of the ecological information contained therein. The photoQuad software is freely available to download and use from: http://www.mar.aegean.gr/sonarlab/photoquad.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Marine benthic habitats constitute an important part of the marine environment, with high ecological, scientific, and economic value (Bianchi and Morri, 2000; Dayton et al., 2000; Hughes et al., 2005); they are composed of a great diversity of sessile benthic species, and provide essential resources to fish and other motile invertebrates (Collie et al., 2000; Thrush and Dayton, 2002). Studies regarding biodiversity assessment, characterization of communities, and monitoring of changes over time commonly involve the estimation of area or coverage of sessile benthic species either through in situ observations or use of underwater photography and video (Bianchi et al., 2004). As the destruction and alteration of benthic habitats increases (Halpern et al., 2007; Pandolfo et al., 2003), so does the need for their accurate and comparable quantitative assessment. The photoQuad software presented here is an image processing tool created to address the need for fast, accurate, and comparative analysis of ecological information contained in digital or digitized photographs, with particular focus on quadrat samples. Quadrat sampling has been widely applied for the assessment of sessile communities in several branches of ecology (Greg-Smith, 1983; Sutherland, 2006), and its main advantage is that sampling is achieved through a non-destructive, relatively fast and cost effective way. With regard to marine benthic habitats, quadrats have been used for the assessment of ecological properties including species composition, abundance and diversity (Bussotti et al., 2006; Martí et al., 2004), spatial patterns (Garrabou et al., 1998; Teixidó et al., 2002), population parameters (Hughes and Jackson, 1985), and for the monitoring of changes due to human and/or naturally induced factors (Garrabou et al., 2002; Roghi et al., 2010; Virgilio et al., 2006). Information that can be acquired from quadrat samples primarily includes species presence/absence observations, direct counts of individuals/colonies to obtain density estimates per unit area, and absolute area or percentage coverage estimates (Bianchi et al., 2004). Among the different methods used to extract such data, three are the most common according to the objectives, the time availability,
and the scale of the study: a) point counts, where points are superimposed on quadrats in a uniform, random or stratified manner, and the number of points assigned to a species sum up to a percentage of the total number of points (Carleton and Done, 1995; Foster, 1991; Leonard and Clark, 1993; Meese and Tomich, 1992); b) grid projection, where a quadrat is broken down into smaller, equally sized grid cells, and the number of cells assigned to a species is interpreted as a percentage of the total number of cells (Benedetti-Cecchi et al., 1996; Bussotti et al., 2006; Dethier et al., 1993; Fraschetti et al., 2001; Virgilio et al., 2006); and c) region-based area or percent coverage estimations, where the absolute or relative cover of a species is ascertained through in situ visual estimates or digitized images (Garrabou et al., 1998; Garrabou et al., 2002; Meese and Tomich, 1992; Pech et al., 2004; Teixidó et al., 2002; Teixidó et al., 2011).

With the advances of digital photography and the introduction of high resolution cameras, still images of quadrats (hereafter referred to as photoc打算fracts) further reduce the time spent on the field, increase the number of samples per survey, facilitate sampling over large spatial and temporal scales or in difficult to approach habitats, and provide permanent records of the communities under study, that can also be used as evidence of ecological conditions to the general public (Collie et al., 2000; Dumas et al., 2009; Lam et al., 2006; Leujak and Ormond, 2007; Preskitt et al., 2004). Still, extraction of results requires laborious post-survey analysis of digital images, and there is a scarcity of dedicated software platforms that facilitate, in an integrated fashion, all major two dimensional (2D) analyses regarding photoquadrat sampling (Nakajima et al., 2010; Pech et al., 2004; Preskitt et al., 2004). Available tools include ImageJ, a public domain image processing application developed by the U.S. National Institutes of Health (http://rsb.info.nih.gov/ij/), Seascape (Teixidó et al., 2011) for areal cover and perimeter estimation based on semi-automatic image segmentation, CPCe (Kohler and Gill, 2006) for random point counts and planar area analysis, PhotoGrid (http://www2.hawaii.edu/~cbird/PhotoGrid/frames.htm) for random point counts, as well as other general purpose image processing software applications.

To our knowledge however, it is common that for the same datasets, but for different core analyses such as point counts, grid projection or region-based species measurements, researchers often need to resort to different software tools, adapting their methods and results’ precision accordingly. This also applies when dealing with extraction of morphometric descriptors or the cross-calibration of different processing methods and the quantitative estimation of their differences. Comparative error analyses among photoquadrat processing methods are thus difficult, and studies usually focus on the assessment of variability among observers (Benedetti-Cecchi et al., 1996), accuracy or precision comparisons between in situ visual methods and a particular photoquadrat processing method (Meese and Tomich, 1992), or the parameterization of a particular method according to target species or study scale (Alquezar and Boyd, 2007; Dumas et al., 2009).

In the effort to resolve these fragmentation issues and assist comparative quantitative studies between various methods and/or datasets, we have developed and validated photoQuad, a new software platform for advanced image processing of 2D photographic samples, dedicated to ecological applications. The system integrates tools for automatic quadrat boundary detection, image calibration, image enhancement, creation of user specific species libraries, random point counts, species counts, estimation of absolute area and relative-to-quadrat coverage using free-hand or segmentation-based regions, and extraction of advanced morphometric measurements. These are functional in a layer-based environment, allowing the efficient processing of quadrat datasets under an integrated format, thereby reducing data redundancy.

![Fig. 1. Screenshot of the photoQuad software, featuring several layer-based analysis objects such as random points, cell counts, species regions, and species markers.](image-url)
The paper describes the main features and functionality of photoQuad, and elaborates on the comparison of grid projection, random point counts, freehand regions, and image segmentation processing methods, in terms of bias in species area and coverage estimation. These are applied on simulated datasets of sessile benthic species, characteristic of the coralligenous communities of the northeastern Mediterranean Sea, with the aim to: a) investigate the variability of estimates across methods, b) examine the effect of the quadrat boundary detection and the subsequent image calibration process on the overall bias, and c) assess photoQuad performance as a new software platform for the analysis of photoquadrat samples.

2. Materials and methods

2.1. photoQuad software overview

photoQuad is a custom software that integrates a suite of 2D analyses used in marine biology and ecology for the study of sessile communities through photographic sampling. It is developed in MATLAB®, but released as a stand-alone application that runs on MS Windows (XP/7) operating systems, and supports several common image formats such as JPG, PNG, TIFF, and BMP. The software has been designed to fully support the direct import and processing of high resolution images (typical specifications: 4000 × 3000 pixels, 24-bit), without the need to manually downsample them prior to analysis. If downsampling is nonetheless needed for some operations due to memory limitations, photoQuad either temporarily performs it internally, or processes the image in blocks, and always delivers back the original sized image.

The photoQuad software operates in a layer-based environment that allows multiple analyses to be performed simultaneously on the same source image (Fig. 1). The advantages of working in layers are the reduction of analysis time, the simplification of data management, and the minimization of data redundancy; results can be reproduced using only the original input image and photoQuad’s native layer file, which can be saved, loaded or modified during processing.

2.1.1. Quadrat detection and image calibration

Outlining the quadrat boundary on a digital image is a common process in photoquadrat processing, used to define the effective sampling area (Aeff). Using image processing operations, photoQuad is able to automatically detect with a single click the inner outline of the quadrat frame against the colored image background. Post-detection corrections can be applied, and a series of tools is available to optionally draw a manual quadrat outline, or select an outline out of predefined adjustable shapes. Image calibration further provides a pixel to real-distance conversion factor (hereafter referred to as the calibration factor, \( f_{\text{cal}} \) pixels × cm\(^{-1}\)), that allows image scaling to metric units (Fig. 2a) and the estimation of actual species area or related descriptors.

An important characteristic of photoQuad is its underlying architecture that enables a flexible, non-linear manner in data exploration and analysis. Analysis objects and their descriptors are dynamically recalculated and referenced to \( A_{\text{eff}} \) each time the quadrat is defined, its outline modified, or image calibration marks are set or altered. In the absence of a defined quadrat, relative-to-quadrat descriptors, such as coverage, are referenced to the whole source image, and effective area \( A_{\text{eff}} \) becomes equal to the whole image area. If analysis objects such as region outlines are defined prior to image calibration, their descriptors are expressed in pixels, and conversion to metric units is automatically performed upon calibration; both unit options are always exported to the final analysis worksheets, plus image calibration metadata. The practical advantage of this architecture is that not only can an image be recalibrated during analysis without work loss, but also that analysis objects can be defined or modified at any particular order, while all recalculations are automatically performed without user intervention.

2.1.2. Species library

The user-defined species library is a core component to photoQuad’s functionality, allowing the association of analysis objects to particular species or group categories. The library can be an MS Excel (*.xls) or a typical comma-separated ASCII file (*.csv), it can be imported or modified at runtime, while custom tools are provided for adding new entries or species preview images during data analysis and exploration.

2.2. photoQuad analysis options

photoQuad integrates a range of 2D photoquadrat analyses (Fig. 2), all of which operate on a distinct layer, can be applied simultaneously, and are further supported by direct ASCII worksheet output. The software does not automatically classify any analysis object to species; this is manually performed by the user.

2.2.1. Species counts (SC)

These are simple markers that flag a particular species or image feature (Fig. 2a), which can be associated to a species library entry; they are used to obtain presence/absence or frequency of occurrence information.

2.2.2. Freehand regions (FH)

The freehand region method refers to the manual drawing of an outline around a specific region of interest (ROI, Fig. 2a), its association to a particular species library entry, and the subsequent extraction of region descriptors. photoQuad is equipped with a variety of tools aiding the interactive selection or modification of ROIs. The method’s basic output descriptors are region area (cm\(^2\)) and coverage (%), which are both

![Fig. 2. Composite figure of the photoQuad software, illustrating different analysis options for a single benthic image; analysis objects are automatically confined into the quadrat’s active area (25 × 25 cm).](image-url)
automatically exported at: a) individual region level, b) species level, grouping all regions assigned to the same species, and c) total image or effective sampling area (Aeff) level. As region area is initially defined by the number of pixels Apix contained within its boundary, once the image is calibrated photoQuad extracts region area in cm² units according to: Aᵢ = Apix × fcb², where fcb is the calibration factor; region coverage is then calculated as Ci = Aᵢ × Aeff / 100. Furthermore, individual region output includes several advanced morphometric 2D descriptors including ROI perimeter Pi (cm), centroid i.e. geometric center coordinates (Xcᵢ, Ycᵢ), or perimeter roughness Ri expressed as the coefficient of variation of the distance in all directions between the region’s boundary and its centroid.

2.2.3. Multi-scale image segmentation regions (SG)

Image segmentation refers to the partitioning of an image into multiple sets of pixels that share some common characteristics such as color or intensity, and Teixidó et al. (2011) demonstrated the effectiveness of segmentation-based methods in seafloor image analysis. photoQuad is equipped with a statistical region merging algorithm developed by Nock and Nielsen (2004), selected upon its efficiency in the successful partitioning of highly complex benthic images. In its current implementation, the source image is sequentially segmented into four different scales from coarse to fine detail (Fig. 2b), facilitating the automatic location of features and boundaries at various scales, and the transformation of the input image into a segmentation map that is easier and faster to analyze (Teixidó et al., 2011), in a user-independent manner. The four different segmentation scales are simultaneously available during photoquad processing. Several image segments can be interactively combined to provide the necessary level of detail required to accurately define and manually classify a particular region of interest (ROI). Each of these ROIs can be subsequently associated to a species library entry and inherit all output descriptors as applied to the FH regions.

2.2.4. Grid cell counts (CL)

The method’s principle is that a grid of square cells with user-defined unit cell area (ACL, cm²) is projected over the image’s effective area Aeff (Fig. 2c). Grid cells that cover particular species or image features can be interactively activated, resulting to a total number of Nᵢ cells per ith species. The output includes area species calculated as Aᵢ = Nᵢ × ACL (cm²), as well as species coverage Ci = Aᵢ × Aeff / 100.

2.2.5. Random point counts (RP)

photoQuad supports three modes of random point generation methods, i.e. random, stratified random, and uniformly distributed (Greig-Smith, 1983). The total number of points(Nrp) can be customized, and points can be spawned either over the whole image, or automatically confined into the quadrat’s active area Aeff (Fig. 2d). Analysis outputs include frequency estimates, determined as the number of points Nᵢ assigned to the ith species/group, or percentage coverage estimates defined as Ci = (Nᵢ × Nrp) / 100. The total number of points assigned to all species per image is also exported, producing an output of total image or total quadrat coverage.

2.3. Study rationale

A comparative analysis for the assessment of accuracy and precision among estimators generally requires a reference dataset of known true values, against which performance is compared. In this context, the accuracy of absolute area or surface coverage estimators cannot be directly compared against field data, given that the true quantities are unknown or even unknowable (Meese and Tomich, 1992). Simulation approaches are feasible alternatives (Dethier et al., 1993; Nadon and Stirling, 2006), and their power is always maximized when their statistical characteristics are backed up by real datasets (Walther and Moore, 2005). In order to compare the different methods (RP, CL, FH, SG) for area and coverage measurements of sessile benthic communities in photoquad samples, simulated data were used as known “real” values, whose statistical properties were derived from empirical distributions of selected organisms meticulously measured from field data samples.

Furthermore, next to the intrinsic bias per measurement method, an additional source of error is the variability in the definition and subsequent calibration of the quadrat’s active area Aeff among different images. This effect is twofold and is illustrated by considering the same region superimposed on two different photograds: (1) variability in the calibration factor fcb among images affects area estimation, due to the different image-distance to real-distance conversion; (2) variability in the definition of the quadrat boundary affects coverage estimation, given that the methods examined estimate coverage as percentage of region or cell area relative to quadrat active area. Random point counts are immune to this source of error as coverage is a function of the number of points only. In order to quantify this error, the same simulated data were overlaid on two sets of photoqgrads: a) the QREF dataset, where all simulated benthic images shared an identical, pre-calibrated quadrat boundary, and b) the QSMC dataset in which both the quadrat boundary and fcb were separately defined per image in realistic time margins. Further errors related to the choice of sampling design or data acquisition in general are not considered in this study.

2.4. Field data

Original data were collected during the summer of 2010, in a study regarding the characterization of coralligenous communities in the Aegean Sea (NE Mediterranean). The study site is located at an offshore reef off Levos island, Greece (N39°19′69.649″, E26°26′175″). High resolution 12.1 megapixel images were taken randomly within coralligenous assemblages at 20–30 m depth, using a still quadrat PVC frame to minimize parallax error, covering an area of 25 × 25 = 625 cm² per image. Images were imported in photoQuad, and a presence/absence assessment was carried out to identify species that were abundant among samples, but with different statistical properties in terms of area and density. Three different species were chosen: the sponge Agelas oroides, a bryozaan Schizomavella sp, and the scleractinian Leptopsammia pruvoti, all of which are considered characteristic of Mediterranean coralligenous communities (Ballesteros, 2006). For each of the aforementioned species, 30 images were drawn out of the available data pool that featured at least one individual occurrence; density Dₛᵢ
(individuals per quadrat) and absolute area $A_{op}$ (cm$^2$) was measured using carefully drawn freehand regions (Table 1), and the respective empirical distributions of density $<D_{op}>$ and area $<A_{op}>$ per species were obtained.

### 2.5. Simulated data

Using custom developed software and these species-specific field data statistics, an equivalent set of simulated patches was created (named as: sp$_1$, sp$_2$, and sp$_3$, Table 1). For each species separately ($A. oroides$, Schizomavella sp., $L. pruvoti$), its empirical density and area distribution was approximated by an exponential $\exp(D_{op})$ and lognormal $\logn(A_{op})$ distribution respectively ($n_{Exp}=10,000$), with corresponding measures of central tendency and dispersion. A density value $d_{op}$ was randomly drawn from $\exp(D_{op})$, and a total number of $d_{op}$ area values were subsequently drawn from $\logn(A_{op})$ in a random manner without replacement. The process was repeated for 30 repetitions per species category, producing disk-shaped simulated patches with accurately known area and density that directly relate to field measurements, and further display a realistic variability both within, and across repetitions. These 30 sets of simulated species patches were superimposed on 30 identical, pre-calibrated quadrats producing the QREF dataset, where each image contained all three simulated species sp$_1$, sp$_2$, sp$_3$. The QIMG dataset was produced by overlaying the identical disk patches on 30 randomly chosen photo-quadrat samples (Fig. 3). In this context, simulation assumptions are: a) simulated species are 2D disk-shaped patches without boundary irregularities; with a gradually fading outline; b) there is no overlapping between species; c) each one of the 30 simulated images per QREF and QIMG datasets contains all 3 species categories; d) the area of each individual patch is considered as the reference “real” area value ($A_{o}$, cm$^2$) in the subsequent analysis.

### 2.6. Data analysis

The simulated datasets were analyzed in photoQuad using the FH, SG, CL, and RP methods, to investigate and compare their bias in terms of accuracy and precision regarding the estimation of absolute area and percentage coverage. Each image of the pre-calibrated QREF dataset ($N=30$) was imported in photoQuad, and all individual patches per simulated species (sp$_1$, sp$_2$, sp$_3$) were marked and measured with the following methods configuration; FH: simulated patch outlines were manually drawn using a maximum zoom level of 200%, without the aid of any semi-automatic outlining tool, SG: images were segmentated and patch outlines were marked at the highest level of segmentation (scale 4 of 4), CL: a unit cell area of 0.5 cm$^2$ was used to define the projected grid that was confined within the quadrat’s active area, RP: a total number of 100 stratified random points were spawned per image, also confined within the quadrat’s active area. For each method and image analyzed, species area and coverage results were exported, and photo-Quad layers were saved on disk. The layer files were subsequently superimposed on the respective QIMG dataset, for which the quadrat was separately outlined and calibrated per image in a deliberately non-meticulous manner, while area and coverage descriptors were automatically recalculated with reference to the new quadrat. Thus, an error analysis on the QREF dataset enabled the investigation of the intrinsic bias per measurement method, while the differences between the QIMG and QREF datasets provided an insight into the additional error associated to the manual recalibration procedure.

Regardless of the dataset and considering each image separately, if $A_{js}$ (cm$^2$) is the total area estimated by method $j$ for species $s$, and $A_{os}$ is the respective known “real” area (cm$^2$), then the unscaled measurement error can be expressed as $\hat{A}_{js}=A_{js}-A_{os}$ (cm$^2$), and the scaled error as $\bar{A}=(A_{js}-A_{os})A_{os}\times 100$ (%). The respective errors $\hat{C}$ and $\bar{C}$ regarding percentage coverage were also extracted. Next to descriptive statistics such as the mean or variance, the mean squared error $\text{MSE}=1/N\times \sum (A_{js}-A_{os})^2$ was calculated, a measure that squares all differences, is sensitive to outliers, but has a different scale than the original measurement (Walther and Moore, 2005). To overcome the unit scaling problem and to provide an alternative measure that is more robust to outliers, the mean absolute error $\text{MAE}=1/N\times \sum |A_{js}-A_{os}|$ was also computed. All measures were extracted both for the scaled and unscaled errors in QREF and QIMG datasets.

### 3. Results

When each measurement method was examined separately, no significant differences were found in absolute area and percentage coverage estimations between the QREF and QIMG datasets at species level (FH: $p_d>0.75$, $p_c>0.95$ for all species; SG: $p_d>0.80$, $p_c>0.97$ for all species; CL: $p_d=1.00$, $p_c>0.71$ for all species; Kruskal–Wallis tests, $p_d$ and $p_c$ refer to area and coverage respectively). The random point (RP) method was not examined here as it is independent of the calibration factor. These results indicate that the image calibration procedure and the redefinition of the quadrat’s boundary has no statistically significant effect on a method’s area or coverage estimation, as far as a still quadrat of fixed dimensions (e.g. 25 x 25 cm) is consistently used, and calibration marks are reasonably placed on image samples.

![Fig. 3. Illustration of the custom simulation software with an example of (a) reference:QREF and (b) recalibrated:QIMG data, where the same set of disk-shaped patches is superimposed on a predefined calibrated quadrat and a random benthic photoquadrat image respectively; patch number denotes species (1: sp$_1$, 2: sp$_2$, 3: sp$_3$).](image-url)
The magnitude of unscaled and scaled errors per measurement method is illustrated in Fig. 4, tabulated by simulated species sp1, sp2, and sp3 that correspond to real species of different density and progressively lower mean area (A. oroids, Schizomavella sp., and L. pruvoti respectively, Table 1). The error among methods shows a consistent pattern across species, both for absolute area and percentage coverage. The segmentation-based method (SG) systematically exhibits the lowest error, followed by the freehand region (FH) method, whereas grid cell counts (CL) and random points (RP) present the poorest overall performance and highest variability across species. The higher variability of CL and RP is particularly pronounced when the scaled errors are examined (Fig. 4b and d), which express the measurement error as a percentage of the species “real” area or coverage per image; the smaller the species, the greater the variability of CL and RP. With respect to under- or overestimation patterns, region-based methods (FH and SG) tend to slightly underestimate area and coverage, while CL and RP generally overestimate the real values. The RP method in particular produces scaled errors up to 300% in species coverage (sp3, L. pruvoti). A notable exception is observed in simulated species sp1 (A. oroids), where the medians of unscaled and scaled RP errors are close to zero, although the method still suffers from high variability in under- or overestimated measurements across images (Fig. 4c,d).

The effect of method type and species category on area and coverage estimates was examined by a nonparametric multivariate analysis of variance (PERMANOVA; Anderson, 2001), using measurement method and species as factors, and measurement error as the response variable (1000 permutations, Table 2). Unless specifically stated otherwise, PERMANOVA results refer to scaled errors. The analysis revealed a significant effect of measurement method on area (F2=312.83, p=0.001) and percentage coverage (F2=51.12, p=0.001) error variability. The effect of species category was also significant (F2=154.46, p=0.001 and F2=53.12, p=0.001), as well as the interaction between both factors (F2=201.16, p=0.001 and F2=23.72, p=0.001).

With respect to the effect of measurement method on area variability, pair-wise t-tests showed statistically significant differences in the multivariate space among all methods (tFH-CL=12.13, tFH-SG=8.35, tSG-CL=7.12, p=0.001). Significant differences were also observed in coverage scaled errors among all measurement methods (tFH-RP=6.39, tFH-CL=8.35, tFH-SG=12.61, tSG-RP=5.93, tSG-CL=7.12, tCL-RP=3.14, p=0.001). However once the unscaled errors were examined, i.e. the direct differences between measured and “real” coverage, CL and RP methods did not differ significantly (tCL-RP=1.37, p=0.132).

Furthermore, pair-wise tests revealed that the statistically significant effect of species category on area error was largely attributed to differences between the smaller species sp2 and the remaining two sp1 and sp3 (t1,2=4.95, t2,3=4.98, p=0.001). The rather comparable size sp1 and sp3 did not differ significantly, particularly in scaled errors (t1,2=0.42, p=0.683). With regard to the effect of species category on coverage estimates, scaled errors of all three species differed significantly (t1,2=2.17, p=0.026, t1,3=6.2, p=0.001, t2,3=5.01, p=0.001), however in their unscaled counterparts, species sp1 and sp2 showed no statistically significant differences (t1,2=0.57, p=0.579).

Next to the investigation of method type and species category contribution to error variability, the actual efficiency of the FH, SG, CL, and RP methods was assessed in terms of their accuracy and precision in species area and coverage estimations. Table 3 presents the unscaled and scaled error analysis per simulated species and measurement method, both for the QREF and QIMG datasets. With regard to accuracy, all measures (mean, MSE, MAE) indicate that the segmentation-based SG method presents the lowest errors across all species and image datasets, both for area and coverage estimates. The sensitive to outliers mean squared error is stable across species,
Table 3
Scaled and unscaled error analysis for area and coverage estimators, tabulated by species and measurement method (simulation results, N=30 photoquadrats). Numbers are rounded to the 2nd decimal, and method abbreviations are FH: freehand regions; SG: segmentation regions; CL: cell counts; RP: random point counts, QREF: reference quadrat dataset, QIMG: recalibrated quadrat dataset. Q25 and Q75 denote the lower and upper quartiles respectively. Refer to main text for definitions of mean squared error (MSE) and mean absolute error (MAE).

<table>
<thead>
<tr>
<th>Species</th>
<th>Method</th>
<th>Area estimation</th>
<th>Mean</th>
<th>Var.</th>
<th>MLE</th>
<th>MAE</th>
<th>Q25</th>
<th>Q75</th>
</tr>
</thead>
<tbody>
<tr>
<td>sp1</td>
<td>QIMG</td>
<td>$\bar{A} - A_0$</td>
<td>1.33</td>
<td>0.80</td>
<td>2.55</td>
<td>1.33</td>
<td>-1.96</td>
<td>-0.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>unscaled error</td>
<td>0.18</td>
<td>0.01</td>
<td>0.04</td>
<td>0.18</td>
<td>0.12</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>scaled error</td>
<td>2.03</td>
<td>0.30</td>
<td>0.35</td>
<td>0.23</td>
<td>0.23</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{A} = (A_j - A_0)$</td>
<td>0.18</td>
<td>0.01</td>
<td>0.04</td>
<td>0.18</td>
<td>0.12</td>
<td>0.26</td>
</tr>
<tr>
<td>sp2</td>
<td>QIMG</td>
<td>$\bar{C} - C_0$</td>
<td>0.89</td>
<td>0.80</td>
<td>2.55</td>
<td>1.33</td>
<td>-1.96</td>
<td>-0.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>unscaled error</td>
<td>0.18</td>
<td>0.01</td>
<td>0.04</td>
<td>0.18</td>
<td>0.12</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>scaled error</td>
<td>2.03</td>
<td>0.30</td>
<td>0.35</td>
<td>0.23</td>
<td>0.23</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{C} = (C_j - C_0)$</td>
<td>0.18</td>
<td>0.01</td>
<td>0.04</td>
<td>0.18</td>
<td>0.12</td>
<td>0.26</td>
</tr>
</tbody>
</table>

is particularly close to zero, and its maximum value is noted in species sp3 (1.46, scaled), which however corresponds to a relative error of only 0.95% and 0.47% in area and coverage respectively, as indicated by the MAE scaled data. Furthermore, error variance is also close to zero across species and datasets, reflecting the high precision of the SG method.

When the FH and CL methods are compared in terms of species area (Table 3), the FH method presents lower unscaled mean errors in the QREF dataset (sp1: -1.33, sp2: -0.35, sp3: -0.36 cm²), with a general tendency to underestimate species area, as indicated by the sign of Q25 and Q75 quartiles. CL has higher respective errors (sp1: 1.81, sp2: 0.39, sp3: 3.18 cm²) and overestimates area. At the QIMG dataset and using area estimates as an example, FH displays higher unscaled mean error (sp1: -2.23, sp2: -0.48, sp3: -0.41 cm²) than the CL method (sp1: 1.81, sp2: 0.39, sp3: 3.18 cm²), but once error is scaled and expressed as a percentage of the “real” values, FH mean error is comparable to CL, but significantly lower for the smaller species sp3 (FH sp3: -7.6, CL sp3: 58.67%). CL errors tend to substantially increase inversely to species size, which is further pronounced at the scaled errors.

In terms of precision, CL method error variance is overall higher than FH, and for both methods, scaled data reveal an inverse relation between variance and species size (FH/QREF sp1: 0.21, sp2: 1.65, sp3: 7.97, CL/QREF sp1: 3.51, sp2: 46.55, sp3: 444.62, scaled data). With regard to coverage, unscaled data show low mean errors and variance, but hide their different sensitivity to species size; their scaled counterparts reveal that the CL method ranks lower to FH, and that its precision drastically deteriorates with decreasing species size. Overall,
results indicate that the freehand region FH method performs better than the grid cell count CL method, both in terms of accuracy and precision.

The random point count RP method presents no differences between the QREF and the QIMG datasets, as it is not sensitive to the image recalibration procedure, except from the fact that the quadrant’s boundary determines the active area in which random points are spawned. Furthermore, area results are not presented in Table 3, as area estimates are inherently not applicable to this method. The scaled values in Table 3 indicate that RP has the greatest error variability compared to all other methods, both in terms of accuracy (mean, MSE, MAE) and precision (variance) measures. Errors drastically increase with decreasing species size (e.g. mean scaled error: sp1: 5.25, sp2: 29.53, sp3: 119.01) and so does the error variance (sp1: 434.27, sp2: 2417.8, sp3: 10,117).

Overall, the error analysis at species level indicates that, regarding the available simulated datasets, the segmentation method (SG) is superior, the freehand region (FH) method ranks second as it shows a lower variability across species compared to grid cell counts (CL), and the random point count (RP) method is accurate on large species, but inherently cannot output absolute area, while its scaled mean absolute errors (MAE) are an order of magnitude greater than all other methods (Table 3). Given that the SG and FH methods have the additional characteristic of outputting measurements at individual patch level, their sensitivity against patch size was examined. Fig. 5 shows coverage scaled errors of the SG and FH methods plotted against individual “real” area A_r and respective coverage C_r, where all species are pooled per dataset, and C_r corresponds to the reference QREF quadrant. Both methods display high mean errors and variance in the estimation of particularly small area patches (A_r ≤ 0.2 cm²), equivalent to C_r ≤ 0.032% quadrant coverage, but the SG method error approximates zero just after the 0.5 cm² A_r area class, with a rapidly decreasing variance. The FH method underestimates the “real” values, and presents a gradual increase in accuracy and precision with increasing patch size. As also demonstrated above, the differences among the reference QREF and the recalibrated Qimg datasets are minimal.

4. Discussion

The photoQuad software presented here was explicitly designed to facilitate fast, standardized, and comparative analysis of digital benthic images, through the integration of multiple processing methods into a single, layer-based platform. The comparative error analysis of four commonly used photoquad measurement methods demonstrates the processing power of photoQuad, and validates its performance on the management and analysis of photoquad samples. On a practical level, multiple measurement methods may not be necessarily applied on the same dataset, but their incorporation into a single software package offers the option to select the most appropriate approach according to the objectives and time limitations of a study, to allow the comparison among methods in terms of measurement bias, or facilitate the cross-calibration of methods should a large number of datasets is available per method. The literature is abundant with studies utilizing different photoquad processing methods such as random point counts (e.g. Carleton and Done, 1995; Dumas et al., 2009; Vroom et al., 2005), grid cell counts (e.g. Benedetti-Cecchi et al., 1996; Fraschetti et al., 2001) or freehand region outlines (Garrabou et al., 1998; Leujak and Ormond, 2007), and photoQuad’s flexibility can facilitate the combination of this fragmented information. Our results further highlight the potential of such software tools to quantitatively assess additional parameters related to photoquad processing, such as the error regarding the repeated quadrat definition and image calibration procedure. As demonstrated through the error analysis, photoquad’s multi-scale segmentation method allows the definition and discrimination of image regions regardless of their size, and the automatic partitioning of high resolution images into segment maps which are easier and faster to analyze. Bernhardt and Griffing (2001) applied a laborious color segmentation approach to estimate benthic species coverage using general purpose image processing software, and suggested that automatic segmentation could not be routinely used for species discrimination, mainly due to limitations of the available algorithm against the high complexity of benthic images. The photoquad software presented here (Fig. 2b), combined with the recent findings of Teixidó et al. (2011) regarding the implementation of a similar algorithm in Seascape, suggests that, when integrated into dedicated software applications, segmentation-based methods are promising tools for benthic image processing.

The experimental approach followed was designed to assess the accuracy and precision of four different measurement methods, and bring to surface their sensitivity to species size, through their application in absolute area and percentage coverage estimation of simulated patches; although a simplified simulation scenario, real data were used to produce realistic density and area patch values. Processing method performance was assessed on the basis of both unscaled and scaled errors, which contain complementary information that should not be overlooked (Walther and Moore, 2005): unscaled error statistics provide a direct expression of performance in absolute units, but they can be misleading with regard to the magnitude of errors relative to the real dimensions of the organisms examined, which is of key importance in the ecological or biological interpretation of results. With regard to the design’s assumptions, although the simulated patches feature statistically comparable area and coverage characteristics to three real benthic species, they constitute a simplified and crude representation of benthic communities. Patches lacked the irregularity in form that is commonly observed in nature, patch colors were uniform per species, and there was no overlapping between patches, as is often the case when dealing with benthic hard substrate communities. These design properties were used in order to reduce the number of variables affecting measurement, and to facilitate comprehension of the basic underlying mechanisms that govern the estimates per method. As it stands however, the design is biased in favor of the segmentation-based region method (SG), which by definition partitions regions according to pixel color statistics. Future studies should increase simulation complexity on a more realistic level, taking into account exhaustive statistical analyses of multiple real species in terms of overlapping, patch size, coloration, and outline irregularity descriptors, as well as the potential correlation of these attributes prior to their use as independent
predictors. Moreover, the number of species per quadrat was only used in order to generate realistic density values for simulated patches, but was not used as an independent variable in the error analysis. Species abundance is a factor known to heavily affect probabilistic methods such as random point counts (Ripley, 2004), but does not affect the free-hand (FH), the segmentation (SG), or the grid cell count (CL) methods. In the context of the simulation approach used here, probabilistic parameters affecting the latter three methods would be related to data sampling design (e.g. number of quadrat samples, size of quadrat frame used, or transect replicates), which were all not considered in this study.

For all measurement methods examined, the image calibration procedure and the redefinition of the quadrat’s outline across image samples was found to be a statistically non-significant factor in the overall photoquadrat analysis process. The importance of this finding mainly reflects upon time efficiency and cross-user variability concerns, suggesting that a meticulous, time consuming quadrat detection and calibration may provide a subjective index of “good” image preparation, but does not necessarily have a statistical impact on the output descriptors. In our experimental design, non-parametric tests regarding area and percentage coverage at species level between the pre-calibrated Q$_{obs}$ and recalibrated Q$_{unc}$ datasets where non-significant for all species categories, and the same applied when results were tested on individual patch or total image level during the preliminary analysis. Of course, the absence of evidence is not evidence of absence, and statistics may have detected differences if a sufficiently large number of test samples were used. Given however the subtle scaled differences between the Q$_{obs}$ and Q$_{unc}$ datasets shown in Fig. 5, or the respective error analysis of Table 3, results should also be considered beyond statistical “significance”. Even when measuring a particularly small individual patch (e.g. < 1 cm$^2$ in Fig. 5), the scaled coverage difference due to the overall calibration procedure was less than 0.5%. Would this difference be significant on the ecological and biological interpretation of the results? Should all other sources of uncertainty in photoquadrat sampling and processing were resolved, the answer is “possibly”. Considering however the wide range of errors involved, including camera lens distortion, image resolution, pixel round-off errors, the measurement of three dimensional species via their two dimensional projection, species overlap, processing method intrinsic errors, and in situ raw data sampling considerations, we would tend to say “no”, but further investigation is required. A sensitivity analysis with different quadrat frame dimensions and more objective measures of user image calibration performance would clarify these questions quantitatively.

The analysis further showed that the bias in absolute area and percentage coverage estimates differed among the four measurement methods, and that methods had different sensitivity to species size. The SG method presented the highest accuracy and lowest error variance, and although favorably biased by the current simulation design, its performance in partitioning highly complex benthic images (Fig. 2b) was also confirmed during the software development process and through preliminary analyses of field data. The method is robust to individual patch size with the exception of extremely small regions (Fig. 5). This effect is related to photoQuadrat’s internal downsampling of high resolution images that is necessary to overcome memory limitations during the computationally demanding segmentation process. Still, the effect is manifested on patches smaller than 0.2 cm$^2$ equivalent to less than 0.032% of quadrat coverage, and the resulting scaled coverage error is barely 2.5%.

With regard to overall performance, the FH method produced fairly low scaled errors at species level that never exceeded 7.6% and 6.8% in area and coverage respectively (Table 3, sp$_2$). It was however less robust to individual size compared to the SG method, a result that agrees with the respective findings of Bernhardt and Griffith (2001). Theoretically, and ignoring pixel round-off errors, the manual drawing of a region would result to a perfect outline of the targeted organism. In real terms however, time limitations restrict the use of a maximum zoom level appropriate for each and every individual organism, which explains why especially the smaller sized patches were measured with lower accuracy (maximum zoom level used: 200%). Moreover, when manually marking a 2D object, it is a human tendency to place region vertices on the object’s outline, thereby producing an inscribed polygon that approximates the actual shape, which eventually underestimates the real area or coverage values.

The grid cell count (CL) method presented here is commonly found in the literature as the “visual method” or “visual estimation” (e.g. Dethier et al., 1993; Fraschetti et al., 2001), but it is important to note that the approaches are equivalent. A larger quadrat is broken down into smaller squares, and percentage coverage is subjectively attributed to an organism according to its contribution in “filling up” each square; photoQuadrat implements the same principle in its digital counterpart, where the option of a custom elementary sampling unit is provided, and the direct estimation of species area is also performed. Furthermore, photoQuadrat’s implementation further increases the method’s accuracy compared to the “visual” approaches, because when an active grid cell falls partially within the quadrat’s active area, the software automatically partitions its contribution and uses the within-quadrat portion of the cell for area and coverage calculations. Using a moderate unit cell area A$_{CL}$ of 0.5 cm$^2$, CL estimates of area and coverage were less accurate than those of FH, and the higher CL error variance showed an inverse relation with species size. In accordance to the results of Dethier et al. (1993), the most apparent differences between CL and the other methods were noted in estimates of simulated species with low “real” coverage. Using a smaller A$_{CL}$ would refine measurements in the expense of processing effort, but would undermine CL’s time efficiency advantage. More importantly, the CL method is generally considered as being prone to subjectivity, but the current experiment did not investigate this parameter; note however that the CL analysis was performed independently by both authors, and the result was a consistent pattern of overestimated measurements (Fig. 4). Dethier et al. (1993) examined user subjectivity in detail, and concluded that the tendency to overestimate abundance of species with low percent coverage was similar across different observers using visual estimates. Benedetti-Cecchi et al. (1996) and Meese and Tomich (1992) also examined observer subjectivity during in situ visual estimates, and discussed the difficulty in reaching conclusive results given the variability in environmental conditions and sampling logistics effects. Although further investigation is required for the digital approach proposed here, initial findings suggest that user response to photoQuadrat’s CL counts is rather comparable.

Random point counts (RP) are widely used for estimating percentage coverage, and their theoretical background and experimental applications are thoroughly discussed by several authors (e.g. Dethier et al., 1993; Greig-Smith, 1983). In this study, RP scaled mean absolute errors were an order of magnitude greater than all other methods, and error variance drastically increased inversely to species size. As shown in Fig. 4d, medians of coverage scaled error regarding the large patches of sp$_1$ (A. oridoides) were close to zero, but the smaller patches of the more abundant sp$_3$ (L. pruvoti) were overestimated by 100%. Being a probabilistic method, repeated RP measurements will eventually converge to the real value (Fig. 4c, sp$_1$), but the number of points required to ensure contact with low coverage species is impractically high. As demonstrated by Dethier et al. (1993), a prohibitively large number of points would be needed to distinguish even moderate differences in percent cover. Nonetheless, when dealing with poorly resolved patches, obscured species boundaries, or poor raw data quality with limited resolution and contrast, point counts can be a useful alternative against region-based methods, as noted by Leujak and Ormond (2007).

Overall, there is a growing body of evidence reporting the deterioration of marine habitats and the decline of biodiversity due to human related activities (Bianchi and Morri, 2000; Coll et al., 2010),
hence research regarding the assessment, mapping, and characterization of benthic habitats and their associated species are vital. Either when focusing on change or on other ecological processes, studies heavily depend upon the representativeness of sampling design, the quality of raw data, the analysis accuracy, and the standardization of processing method protocols, particularly when biodiversity changes are examined over different spatiotemporal scales (Roghi et al., 2010; Virgilio et al., 2006), or when data are combined for meta-analyses (Côté et al., 2005). The available literature reveals a labors research effort on the assessment of marine benthic communities, which however is fragmented into several processing methods with different parameterization, thereby undermining the ability to effectively compare results among studies on a wide scale, and to assess the amount of variance in their compilation. The photoQuaD software is a powerful tool for the analysis of digital benthic images, and the flexible integration of the most commonly used processing methods under a unified format aims to help resolve these discrepancies, at least to the amount that they arise due to the raw data processing tools and methods implementation.

The photoQuaD software is freely available to download and use from: http://www.mar.aegean.gr/sonarlab/photoquad.

Acknowledgments

We thank Prof. Drossos Koutsoubs and Vasilis Gerouvelis for their support and discussions regarding the improvement of photoQuaD’s functionality. We also thank Zacharias Kapelonis for his comments on the manuscript. We greatly appreciate the specific support of Núria Teixidó and colleagues (Institut Ciències del Mar, ICM-CSIC), who provided valuable insights into the Seascape software and the application of image segmentation to benthic images, as well as critical comments on the manuscript. The Editor S.E. Shumway and two anonymous referees are thanked for their constructive reviews. [SS]

References


