Methods in Oceanography 17 (2016) 169–186

Contents lists available at ScienceDirect

Methods in Oceanography

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

Full length article

Combining imperfect automated annotations of underwater images with human annotations to obtain precise and unbiased population estimates

Jui-Han Chang[a], Deborah R. Hart[a,*], Burton V. Shank[a], Scott M. Gallager[b], Peter Honig[b], Amber D. York[b]

[a] *Northeast Fisheries Science Center, National Marine Fisheries Service, 166 Water St., Woods Hole, MA 02543, USA*

[b] *Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA*

HIGHLIGHTS

- Methods for estimating population size from imperfect image data are analyzed.
- Stratification with a local ratio or regression estimator is generally most effective.
- The false negative rate is key in determining the best method.

ARTICLE INFO

Article history:
Received 2 September 2015
Received in revised form
14 August 2016
Accepted 19 September 2016

Keywords:
Underwater imagery
Computer vision
Population estimation
Scallop
Geographically weighted regression

ABSTRACT

Optical methods for surveying populations are becoming increasingly popular. These methods often produce hundreds of thousands to millions of images, making it impractical to analyze all the images manually by human annotators. Computer vision software can rapidly annotate these images, but their error rates are often substantial, vary spatially and are autocorrelated. Hence, population estimates based on the raw computer automated counts can be seriously biased. We evaluated four estimators that combine automated annotations of all the images with manual annotations from a random sample to obtain (approximately) unbiased population estimates, namely: ratio, offset, and linear regression estimators as well as the mean of the manual annotations only. Each of these estimators was applied either globally or locally (i.e., either all data were used or only those near the point in question, to take into

* Corresponding author.
E-mail address: deborah.hart@noaa.gov (D.R. Hart).

http://dx.doi.org/10.1016/j.mio.2016.09.006
2211-1220/Published by Elsevier B.V.

account spatial variability and autocorrelation in error rates). We also investigated a simple stratification scheme that splits the images into two strata, based on whether the automated annotator detected no targets or at least one target. The 16 methods resulting from a combination of four estimators, global or local estimation, and one stratum or two strata, were evaluated using simulations and field data. Our results indicated that the probability of a false negative is the key factor determining the best method, regardless of the probability of false positives. Stratification was the most effective method in improving the accuracy and precision of the estimates, provided the false negative rate was not too high. If the probability of false negatives is low, stratified estimation with the local ratio estimator or local regression (essentially geographically weighted regression) is best. If the probability of false negatives is high, no stratification with a simple global linear regression or simply the manual sample mean alone is recommended.

1. Introduction

Underwater optical surveys of fish and invertebrate populations are becoming increasingly common (e.g., Davis et al., 1992; Gallager et al., 2005; Howland et al., 2006; Yoklavich et al., 2007; Rosenkranz et al., 2008; Taylor et al., 2008; Tolimieri et al., 2008; Singh et al., 2013 and Gallager et al., 2014). Such surveys have numerous advantages over traditional surveys using fishing gear, including being able to observe populations at all scales under natural conditions, and detection efficiency that potentially approaches 100%. Optical surveys often generate hundreds of thousands to millions of images. Manually annotating all of the images (i.e., having people identifying the targets of interest in each image) would thus often be impractical. The traditional statistical approach to this problem would be to only manually annotate a sample of the images and obtain inferences on the population (which for our purposes is defined as the targets contained in all of the collected images) based on the sample. Alternatively, computer vision software can produce "automated annotations" that identify the targets in every image. However, automated annotators can make errors, both because they may not detect some targets ("false negatives") and because the annotator mistakenly identifies some objects ("distractors") as targets when they are not ("false positives"). Thus, analyses based on the raw automated counts can be seriously biased. Errors from automated annotations are often autocorrelated and spatially non-stationary due to, for example, a certain region having high densities of distractors or reduced visibility. Manual annotations of a sample of the images can help detect and correct for errors by the automated annotators, in which case the goal is to produce estimators for the population, based on the combination of automated and manual annotations that are more efficient than using the manual annotations alone (i.e., the variances of estimators are less than the variance of the sample mean of the manual images), as well as being at least approximately unbiased.

Although there have been numerous studies devoted to automated detection and classification of marine organisms (e.g., Culverhouse et al., 2006; Marcos et al., 2008; Spampinato et al., 2008 and Beijbom et al., 2012), these studies usually conclude with estimating confusion matrices or error rates. The final task of obtaining estimates of the population of targets in all images from automated annotations that contain errors has received less attention. Solow et al. (2001) considered the situation where classification of plankton samples may be in error, which was corrected by inverting the confusion matrix (see also Hu and Davis, 2006 and Verikas et al., 2015). The problem they considered is simpler than the one we are considering here because they were only concerned with classification of an object but not its detection, and because errors were assumed to be stationary and