SC/67A/PH/02

Happywhale: Globalizing Marine Mammal Photo Identification via a Citizen Science Web Platform

Ted Cheeseman, Tory Johnson, Ken Southerland, Noah Muldavin

INTERNATIONAL WHALING COMMISSION

Papers submitted to the IWC are produced to advance discussions within that meeting; they may be preliminary or exploratory.
It is important that if you wish to cite this paper outside the context of an IWC meeting, you notify the author at least six weeks before it is cited to ensure that it has not been superseded or found to contain errors.

Happywhale: Globalizing Marine Mammal Photo Identification via a Citizen Science Web Platform

Ted Cheeseman, Tory Johnson, Ken Southerland and Noah Muldavin
Happywhale.com, 904 Columbia St, Santa Cruz, California, USA

Abstract

Happywhale is a web-based marine mammal photo ID crowd-sourcing platform that as of April 2017 has been online at Happywhale.com for 20 months. We remain very much in active system development, guided by user feedback in pursuit of our dual, complementary goals of engaging citizen scientists and using that engagement to generate high quality, low cost marine mammal photo ID data together with web-based tools of value to marine mammal scientists. Since inception we have received submissions of over 41,000 images contributed by over 1000 scientists and citizen scientists. Within these images we have recorded 32 cetacean species while focusing individual ID efforts on humpback whales in collaboration primarily with Cascadia Research Collective, Allied Whale and Alaska Whale Foundation. The site currently displays encounters of 4813 individuals in 10124 encounters. We have focused development efforts on image management efficiency, with an implementation of an individual ID image recognition algorithm for humpback fluke matching now available to collaborating research groups. Efficient image management and ID allowed us to find long-distance matches between image catalogs that would not otherwise have been compared, and enabled us to contribute to entanglement response efforts by identifying whales along the California coast. We anticipate broadening collaboration, especially across the northwest Pacific where the dataset for humpback whales is achieving sufficient coverage to return match rates of approximately 5 to 70% depending on region and image quality. We believe we are now enabling ongoing ocean-basin wide population study at a scale and economy not previously feasible without web-based collaboration and automation.

Introduction

In 2016 we reported to IWC on the first season of our web based marine mammal photo ID platform Happywhale (Cheeseman and Southerland 2016). This paper is intended as an update of efforts, developments and results since that time.

Methods

We released Happywhale.com v1.0 in August 2015 and continue to be primarily focused on system development and managing a steadily increasing flow of data. Development priorities are directed by user feedback in pursuit of our dual goals of public engagement and scientific contribution. We continue to follow image management processes as reported to IWC in Cheeseman and Southerland (2016), with substantial improvements to capacity and efficiency, and to user experience. Of note has been the implementation of "id.happywhale", an interface to an automated image recognition algorithm trained on north Pacific humpback whale flukes (algorithm adapted from Stewart et al, unpublished) that allows significant efficiency improvements when processing images. Because of the unpublished status of the current algorithm, we undertook only a cursory accuracy assessment with 186 known matched flukes to test its usefulness and understand its behavior. With these developments we are able to process images relatively rapidly and partially reconcile previously unmatched datasets. We thus began to think of humpback whale ID efforts on a global scale, expanding beyond our initial two focal areas of coastal California and the Antarctic Peninsula. We began a collaboration with Alaska Whale Foundation, adding 340 individual humpbacks provided by Fred Sharpe to our reference catalog, along with images collected worldwide from conservation and outreach groups, tour operators and individuals.

Each image or set of images is processed to optimize potential automated matching success with cropping, rotation and exposure adjustment as necessary, then matched against existing known whales.

Where images were received with individual ID naming and/or numbering from the contributor, these IDs are preserved alongside any newly found match ID information. Unmatched whales with flukes of sufficient quality such that we are confident we could match the whale if resighted are incorporated into the reference catalog once we have manually confirmed the individual is not represented in the regional catalog, using NOAA Distinct Population Segment regions (figure 1). We do not attempt to manually confirm that unmatched whales are unique between regions, only within regions. Manually confirming an individual is new to the catalog is the most time intensive step in catalog management, inherently leading to a backlog of unmatched flukes. Immediately following each periodic update to our reference catalog, all unmatched images are re-tested through id.happywhale.com, attempting to match several thousand images typically in a matter of a few hours.

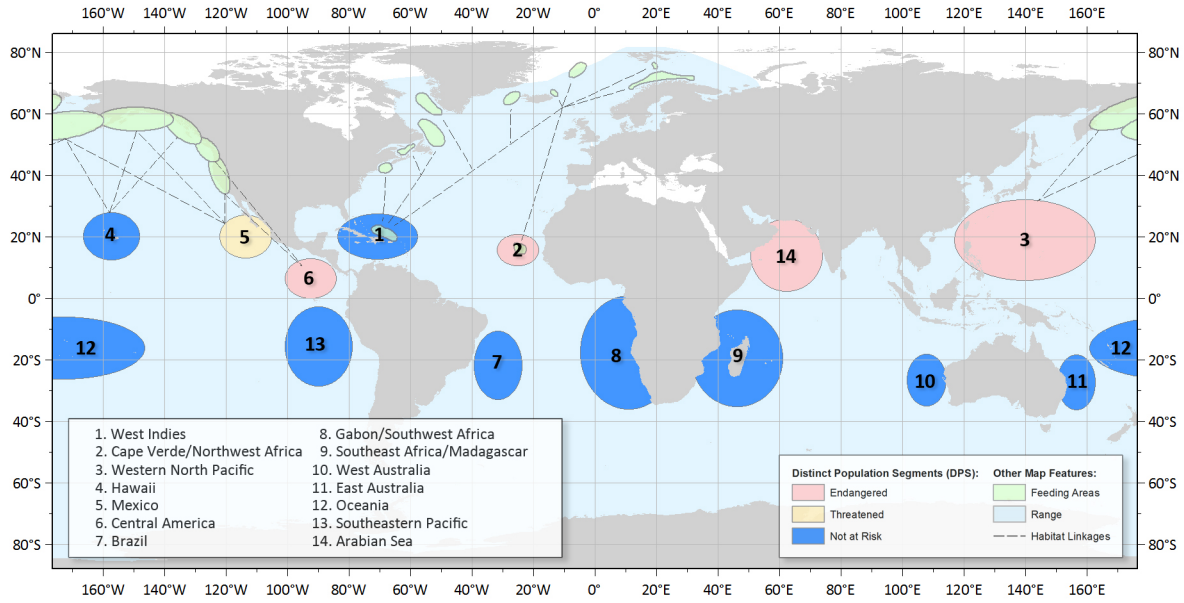

Figure 1 Humpback Whale Distinct Population Segments (NOAA 2017)

By means of an automated notification system, all image contributors are notified of any identifications made, matches found, and repeat sightings, with links to the data online, visually represented but without accessible sighting detail such as map coordinates or precision. For research collaborators, we provide access to the curated encounter data for matches found to their whales and whales encountered in their regions of interest. Collaborating researchers are free to use data sourced from Happywhale with the single requirement that they respect Creative Commons usage rights[1] as designated by data contributors. Most users do not change settings from the default of Public Domain (i.e., open access) rights.

With the capability of rapid individual identification, we also now respond during entanglement incidents when ID images are provided by California Whale Rescue or others of whales involved in entanglement, ship strike or stranding incidents.

Results

Since inception in August 2015 we have received submissions of over 41,000 images contributed by over 1000 scientists and citizen scientists. Within these images we have recorded encounters of 32 cetacean species. We cataloged all encounters and provided data to researchers where the interest has been expressed to us (Table 1).

[1] See https://en.wikipedia.org/wiki/Creative_Commons_license for details on Creative Commons usage rights

Table 1 Partial list of cetacean encounter data recipients from Happywhale

Researcher/Research group	Species of interest	Region of interest
Alaska Whale Foundation	Humpback Whales	Southeast Alaska
Allied Whale	Humpback Whales	Southern Hemisphere and North Atlantic
California Whale Rescue	Humpback Whales (entanglements)	California
Caroline Weir / Falklands Conservation	Sei Whales	South Atlantic especially Falkland Islands
Cascadia Research Collective	Humpback Whales	west coast Central America through British Columbia
Jorge Urban	Humpback Whales	West coast Mexico
Mariano Sironi and Vicky Rowntree	Southern Right Whales	South Atlantic / Antarctica
Marine Mammals of Oaxaca	Humpback Whales	Matches to Oaxaca humpbacks
Mingan Island Cetacean Study	Blue Whales	North Atlantic
Panacetacea	Humpback Whales	Matches to Panama humpbacks
Paula Olson	Antarctic Blue Whales	Antarctica
Robert Pitman	Antarctic Killer Whales	Antarctica
Whales of Guerrero Resarch Project	Humpback Whales	Matches to Guerrero, Mexico humpbacks

Our test of known matched images with id.happywhale using an adaptation of an open-source draft image recognition algorithm (Stewart et al, unpublished) yielded match accuracies of 70% overall (Table 2). We found id.happywhale sufficiently effective to reduce per image matching effort by approximately two thirds when processing a set of images completely via automated and manual confirmation of presence or absence in a reference set. And we found id.happywhale to reduce large-batch image matching attempts to near zero effort when not following automated matching with manual searching for unmatched whales. This has proven efficient enough to undertake long range matching efforts, yielding long range matches of interest such as seven between the Cascadia US West Coast dataset and whales known from Alaska, five between Baja California, Mexico and Alaska, and one match from Hawaii to California.

Table 2 Preliminary accuracy assessment of id.happywhale image recognition for humpback fluke matching

pigment category	accuracy	N=
1 (0 to 10% black)	70.00%	10
2 (10 to 35% black)	90.48%	21
3 (35 to 65% black)	77.42%	31
4 (65 to 90% black)	82.61%	46
5 (90 to 100% black)	55.13%	78
Cumulative	70.43%	186

As of April 2017 we have collected data on 4813 individuals humpback in 10,124 identified encounters (Figure 2). This continues to be concentrated in our original focal areas of coastal California (6671 identified encounters of 2470 individuals) and the Antarctic Peninsula (937 identified encounters of 865 individuals), though is increasingly spread especially north and south along the west coast of North America, and from Hawaii. In Antarctica we have been collaborating with Allied Whale since Happywhale's inception; this collaboration has expanded with Allied Whale testing id.happywhale for possible use in their internal matching methodology, testing that to date has yielded matches to 92

individuals known from the North Atlantic Humpback Whale Catalog, spread geographically from the Silver Bank of the Dominican Republic to the Azores to Iceland, Greenland and Svalbard.

Figure 2 Distribution of identified humpback whale encounters in Happywhale as of April 2017

Finally, working with California Whale Rescue to provide rapid feedback for entanglement response in the rare cases where an entangled whale flukes up and is photographed, we were able to identify individuals in three cases during entanglements, and in one separate case were able to provide resight data for a whale seen in apparent good health six months after a successful disentanglement. We have since implemented specific image tagging and encounter attribute functionality to better enable tracking of known formerly entangled or otherwise human-impacted whales.

Discussion

Last year we reported what we felt was a successful pilot year for Happywhale where our efforts showed promise both in generating data of value and creating meaningful public engagement in through citizen science (Cheeseman and Southerland 2016). We feel that the platform has matured substantially in the year since. The rewards have been (1) high identification rates for whales in waters where we have access to extensive reference catalogs (principally California, and to a lesser but increasing extent, parts of SE Alaska and the Washington/British Columbia border region, (2) efficient creation and effective sharing of meaningful data to Cascadia Research Collective, Allied Whale, Alaska Whale Foundation, California Whale Rescue, and several other research and conservation organizations (partial list of collaborating organizations, and major contributing companies and individuals below), (3) substantial interest from multiple research groups in developing collaborations particularly in northwest Pacific Ocean waters, and (4) encouraging and passionate engagement from the public, in particular from whale watch tourism naturalists. A favorite recent quote, “Thanks to [Happywhale] we are learning more and more, empowering citizen scientists in the most effective way possible.” ~ Ryan Lawler, Newport Coastal Adventure, April 2017.

While we recognize the datasets are categorically very different, we are encouraged by the potential shown with over 10,000 identified encounters to date, considering that the SPLASH study was built on 8000 encounters. We intend to continue to develop both to improve the value and capacity of research tools and the user experience to broaden and deepen user engagement. In keeping with these missions, we

welcome feedback and collaboration with scientists for whom the tools and/or data developed by Happywhale.

Partial list of scientific collaborators sharing encounter data[2]:
Alaska Whale Foundation
Allied Whale
Cascadia Research Collective
Marine Mammals of Oaxaca
Niue Whale Research Project
Panacetacea
Whales of Guerrero Research Project

Major contributing tour operators (with more than 100 contributed encounters)
Blue Ocean Whale Watch
CaboTrek
Cheesemans' Ecology Safaris
Fast Raft Ocean Safaris
G Adventures
Grand Circle Expeditions
La Orca de Sayulita
Lindblad Expeditions
Monterey Bay Whale Watch
Newport Coastal Adventures
Newport Whales
Oceanwide Expeditions
Polar Latitudes
Quark Expeditions

Major individual contributors (with more than 100 contributed encounters)
Mark Cornish
Jodi Frediani
Alethea Leddy
Scott Portelli / Tongan Fluke Collective
Denny Zwiefelhofer

Acknowledgements

We wish to thank John Calambokidis and Kiirsten Flynn at CRC, Peter Stevick and Madeline Kellett at Allied Whale, College of the Atlantic, and Fred Sharpe of Alaska Whale Foundation for support with image catalog access.

References

Cheeseman, T. and Sutherland, K., 2016. Report to International Whaling Commission *SC/66b/SH*

NOAA National Marine Fisheries Service. Humpback Whale *(Megaptera novaeangliae)* Protected Resource Information Page. http://www.nmfs.noaa.gov/pr/species/mammals/whales/humpback-whale.html. Last updated January 2017. Accessed April 2017.

[2] With some collaborators, only partial datasets have been shared with us, depending on the nature of data, previous sharing agreements, and/or other factors

Stewart, C., Berger-Wolf, T., Holmberg, J., Van Oast, J., unpublished open source code of IBEIS automated image recognition algorithm distributed under GPL v2 open source license accessed March 2016. See IBEIS.org.