Using Wild I.D. as a reliable source for mark and recapture studies on northern pike (*Esox lucius*)

Martin Evans
Aquatic Biology Program
Bemidji State University

Mark and recapture studies are a very popular method fisheries biologists use to assess certain fish populations in lakes. This process can be very labor intensive and expensive. Wild I.D. is free software developed by Dartmouth College that uses SIFT program to find unique features in photographs. Initially developed for identification of African land mammals the program gives each photo a score and percent match to other photos. Northern pike were used in this study to determine if the program can recognize simulated recapture events. Photos of sample fish were taken at two separate locations, the photos were then copied and cropped four different ways: Original Full, Original Middle, Negative Full, and Negative Middle. The program had a median percent match of 0.84, 0.85, 0.84, and 0.87, respectively. A Kruskal-Wallis test was ran and a non-significant difference was found between the four sample groups (p-value = 0.90). For all 120 of our trials we found that the simulated recapture was the program's top rated choice for being the best possible match. That gave the program a 100% success rating in choosing the recaptured fish. We are unaware if the software is able to identify a fish once the markings get distorted as the fish grows. Future studies will need to determine if there is any cause for concern about this possibility.

Faculty Sponsor: Dr. Andrew W. Hafs

Introduction

In a mark and recapture study, fish are captured, marked with one of a variety of tags, released, and re-captured many times by repeated sampling. This type of study allows biologists to look at the individual growth rate of a fish, estimate abundance, and look at the fish population in real time. However, this type of study can be expensive and very time consuming.

A more cost effective and less labor-intensive method could change how mark recapture studies are performed and would increase the use of these valuable studies. Wild I.D. is a photo recognition program that uses SIFT (scale invariant feature transform) technology to locate and pick out specific features in a photo (Bolger 2015).

Photo recognition can be a valuable and cost effective tool for biologists to use in fish population assessments (Meekan and Bradshaw 2006). Photo-identification of individual dolphins and whales has been an important part of research on the behavior and biology of cetaceans since the early 1970's (Würsig and Würsig 1977). Hillman et al. (2003) found by using computer assisted identification the time needed to identify several species of whales and dolphins was cut down from 19 minutes, manually, to 5 minutes, with a computer.

Testing the Wild I.D. program to make sure it is accurate is the main focus of this study. The secondary focus of this study is to determine if using the middle third of northern pike (*Esox lucius*) above the pelvic fin is as accurate as using the entire fish. The third focus of this study is to determine if altering the photographs, turning them to a negative, will increase or decrease the accuracy of the program.

Methods

Pictures of northern pike were taken during two separate events at two separate locations. A total of 59 pictures were taken at the Tamarack River located in Waskish, MN in March 2015. These pictures were taken while helping with a project between the Minnesota Department of Natural Resources and Bemidji State University. Pictures were taken during the spring Fyke netting assessment for northern pike. Once captured in Fyke nets, the fish were put onto a measuring board to have length and sex recorded. The fish was oriented with head to the left and ventral side of fish at the

bottom of the picture. Once measured, a picture of the fish was taken using an iPhone (iPhone 6; Apple; 8 megapixel camera; 29 mm focal length) with the intent to capture as much of the fish as possible.

A total of 33 pictures of northern pike were taken at the second site while assisting in a tagging and recapture study conducted by the United States Geological Survey, Minnesota Department of Natural Resources, and private landowners on Shingobee Lake located in Akeley, MN in March 2015. These northern pike were also captured using Fyke nets. Once captured, the fish were put onto a measuring board to have length and sex recorded. The fish were visually scanned for Floy tags placed from previous captures. If no tag was present a new tag provided from the Minnesota Department of Natural Resources was placed in the fish. In addition to the DNR tag, a tag provided from Bemidji State University Aquatic Biology program was placed in each northern pike captured. The new tags were to be used for future recapture data. If already tagged, the number on the tag was recorded in addition to the length and sex of the fish. If a tag was not present prior to capture, a tag was added and the number was recorded for future reference. Once measured, a picture of the fish was taken, using the same iPhone as stated before, with the intent to capture as much of the fish as possible with the same orientation as the Tamarack sample.

Edits on the pictures were performed in the application *Paint* in Microsoft Windows 10. The pictures were cropped in two ways. The first was to try and get as much of the fish as possible (Figure 1). The second was to only use the middle third of the fish above the pelvic fin (Figure 2). Once the pictures were cropped the color was inverted or they were turned to a negative (Figure 3) to test to see if it made a significate difference when using the program. All pictures were oriented with the head of the fish to the left and the belly of the fish to the bottom of the picture. In addition to these cropping guidelines, all pictures were cropped to remove the background or items that may be covering the fish color pattern in the picture.

The Wild I.D. program uses the unique features in each photo to match and score them against other photos uploaded into the program to detect for matches. Each photo receives a percent match score, which is the likelihood that this fish matches the fish in the other pictures. The program uses scale-invariant features from each picture to compare against. This allows the pictures being compared to be of different scale and rotation (Lowe 2004).

After all of a pictures were cropped, a random number generator was used to determine which picture would be used as the recapture event for each trial. The picture was then re-cropped and added to the folder with the pictures to be ran through the program. The pictures were run through the Wild I.D. program following the user instructions that are provided in the program download. Thirty scoring trials were run for each of the picture sets. Once the scoring was complete the user would click through all of the pictures and either select or reject a match. For each trial a file was created containing the picture I.D. name, the ranking out of twenty the program gave the recapture, and the percent match between the base picture and the recapture.

Program R was used to run a Shapiro-Wilk test to determine if the data had a normal distribution (Shapiro and Wilk 1965). Because it was non-normal, a Kruskal–Wallis test was then used to test for statistical differences in the percent match between the four sample groups (Kruskal and Wallis 1952).

Results

For all 120 of our trials the simulated recapture was the program's top rated choice for being the best possible match. That gave the program a 100% success rating in choosing the recapture fish. There was not a significant difference in the percent match between the four groups (p-value = 0.90). The medians and interquartile ranges of the percent matches from the four trial groups can be found in Table 1.

Figure 1. - Example of original full portion of northern pike after it has been cropped from the original photograph taken in the field.

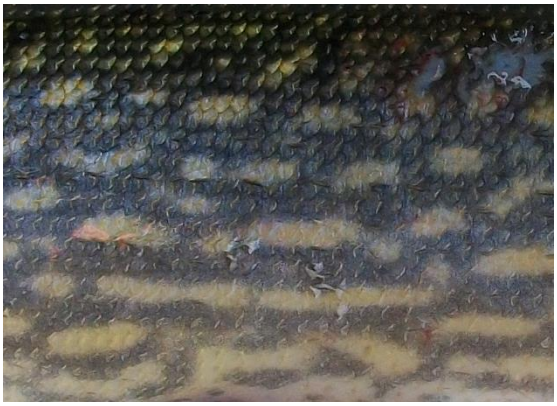

Figure 2. - Example of original middle third portion above the pelvic fin of northern pike after it has been cropped from the original photograph taken in the field.

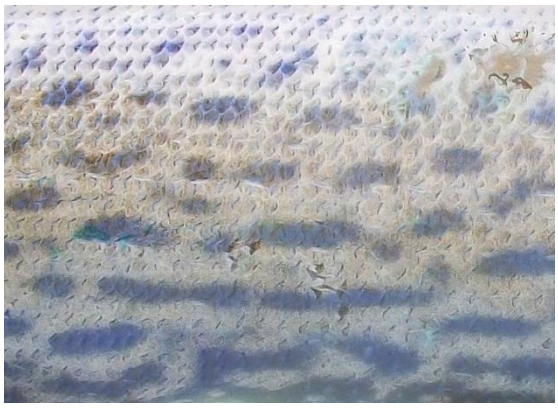

Figure 3. - Example of negative middle third portion above the pelvic fin of northern pike after it has been cropped from the original photograph taken in the field and had the color inverted.

Table 1. - Medians and interquartile ranges (IQR) of the percent matches for the four different sample groups: Original Full, Original Middle, Negative Full, and Negative Middle.

	Original Full	Original Middle	Negative Full	Negative Middle
Median	0.84	0.85	0.84	0.87
IQR	0.14	0.10	0.18	0.12

Discussion

With little information available on photo recognition of fish, the verification of the Wild I.D. program is a beginning step in the process of finding a reliable and cost effective way to identify an individual fish. The majority of photo recognition testing has been performed on terrestrial species (Bradshaw et al. 2007) with the exception of a few marine species of high importance including dolphins and whales (Würsig and Würsig 1977). Photographic identification of cetaceans is a powerful and relatively benign technique which, at least for dolphins, has not yet reached full potential (Würsig and Jefferson 1990).

Trail camera recapture studies on mammals have shown that identification of individual animals of a species is most accurately done on species with distinct color patters (Trolle and Kery 2003). Our results provide evidence to support the same holds true with fish identification. This could possibly limit the number of fish species that the software can be used on because not all fish species have distinct color patterns like northern pike.

The Wild I.D. program did not find that changing the crop location or inverting the color of the photo made a significant difference between groups on the basis of percent match, however it had a 100% success rate when picking the "recaptured" fish for all 120 trials of northern pike. This shows the program is reliable, however, future studies will be the true test of the program. We are unaware if the software is able to identify a fish once the markings get distorted as the fish grows. A possible solution to this problem would be to test which markings or features of a fish stay more consistent throughout their life. For example, the patterns on the cheek of a fish may be more reliable than the coloration pattern from the middle third of the fish. Once the program is thoroughly tested, we believe it could open up endless possibilities and become a well-used tool among fisheries biologists in fisheries studies.

Acknowledgements

I would like to thank Jon Brill, Jake Graham, Chester Powell, Brad Anderson, United States Geological Survey, and the Minnesota Department of Natural Resources.

References

Bradshaw, C., C. Speed., and M. Meekan. 2007. Spot the match-wildlife photo-identification using information theory. Frontiers in Zoology 4:1-11.

Bolger, D. 2015. Wild I.D. The program can be downloaded at http://dartmouth.edu/faculty-directory/douglas-thomas-bolger.

Hillman, G.R., B. Würsig., A. Gailey., N. Kehtarnavaz., A. Drobyshevsky., B.N. Araabi., H.D. Tagare, and D.W. Weller. 2003. Computer-assisted photo-identification of individual marine vertebrates: a multi-species system. Aquatic Mammals 29.1:117-123.

Hollander, M., and D.A. Wolfe. 1973. Nonparametric Statistical Methods. John Wiley and Sons.

Kruskal, W., and W.A. Wallis. 1952. Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association 47:583-621.

Lowe, D. 2004. Distinctive image features from scale-invariant key points. International Journal of Computer Vision 60:91-110.

Meekan, M., and C. Bradshaw. 2006. Population size and structure of whale sharks Rhincodon types at Ningaloo Reef, Western Australia. Marine Ecology Progress Series 319:275-285.

R Core Team 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Shapiro, S., and M. Wilk. 1965. An analysis of variance test for normality. Biometrika 52:591-611.

Trolle, M., and M. Kery. 2003. Estimation of ocelot density in the Pantanal using capture-recapture analysis of camera trapping data. Journal of Mammalogy 84:607-614.

Würsig, B., and T.A. Jefferson. 1990. Methods of photo-identification for small cetaceans. Report of the International Whaling Commission 12:43-52.

Würsig, B., and M. Würsig. 1977. The photographic determination of group size, composition, and stability of coastal porpoises (*Tursiops truncates*). Science 198:755-756.