# SUMMARY OF PHOTO-IDENTIFICATION INTERNSHIP AT THE SCOTTISH ASSOCIATION FOR MARINE SCIENCE

This report reviews the photo-identification internship at the Scottish Association for Marine Science (SAMS) with Dr Steven Benjamins in conjunction with the SIORC Community Project as part of the Common Skate Photographic Identification Database of Scotland (COSPIDS) project. It details the work carried out during the internship, the skills developed and the output of the internship. It was funded through the MASTS Marine Biodiversity, Function and services theme.

## PATTERN-RECOGNITION SOFTWARE TRIALS

The primary focus of the internship was to develop skills in photo-identification. Such skills included learning how software packages vary in the functionality and capabilities, how to use the different software programs and learning about the implications and complications of photo-identification. The initial task was to test and compare different pattern-recognition softwares. After preliminary research into the different software programs available, three software packages – Extract Compare, Interactive Individual Identification System (i3S) and HotSpotter - were selected to be tested. Demos of the softwares were used to learn the basics of the software before using a small group of sample images to test the software. The following sections give a brief overview of the three softwares trialled.

### EXTRACT COMPARE

Extract Compare is a free pattern matching software developed by Conservation Research Ltd. The software works on the basis that submitted photos are focused on the same region of an organism and taken from the same perspective. Guided by the user highlighting the outline and dorsal midline of the organism (see Figure 1), ExtractCompare creates a three-dimensional mesh around the organism in the image, as shown in Figure 2. The mesh can be fitted and refitted multiple times, increasing the accuracy of the mesh around the points each time, as illustrated by the comparison between figure 2 and 3. The software then extracts the pattern from the region of interest and based on information from the photo and the

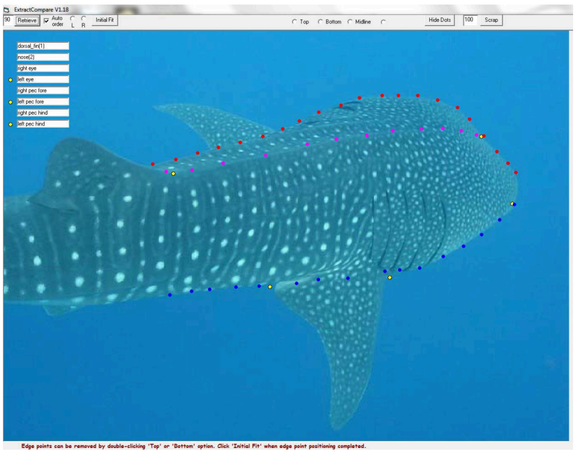

*Figure 1. Screengrab of ExtractCompare after placing the "special points". Blue dots outline the right-hand side of the animal, red dots the left-hand side, magenta dots show the dorsal midline and yellow dots show the key features.*

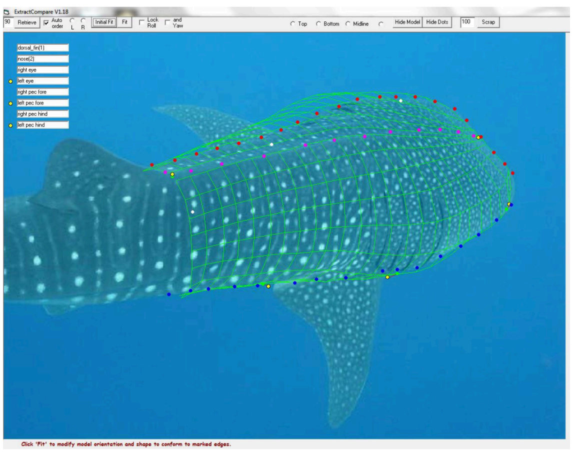

*Figure 2. Screengrab of the initial mesh fitting around the image based on the spots outlining the organism.*

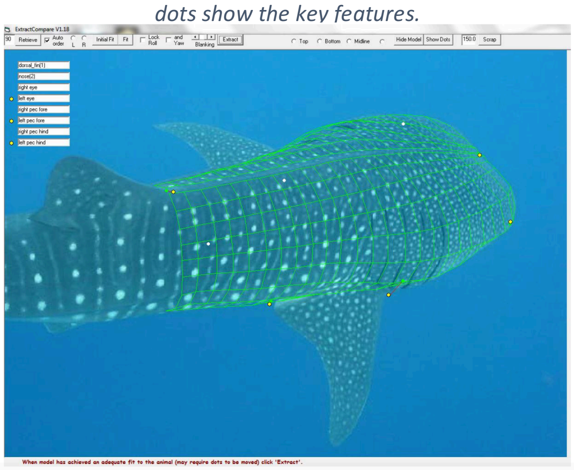

*Figure 3. Screengrab of the mesh after being refitted multiple times around the image.*

Danielle Sloan University of Glasgow danielle_sloan@aol.co.uk

data points within the mesh model (which compensate for the orientation and shape of the body) creates a pattern cell. In the final step, ExtractCompare processes and compares the pattern cell to those already present in the database, producing a ranking of similarity between new and existing pattern cells. Despite the uniqueness of ExtractCompare's approach to pattern-matching, the software was deemed unsuitable for the COSPIDS project due to complexity of the database required to run the software. As a result of the database complexity the sample images were not tested in ExtractCompare. Furthermore, the lack of available IT support from Conservation Research Ltd. at the time meant that the software was not a viable option at that point in time.

## INTERACTIVE INDIVDIUAL IDENTIFICATION SYSTEM (i3S)

I3S is a free photo-identification application which uses natural markings to identify individuals. i3S requires the user to manually select the organism's distinguishing features - including reference features and spot patterns – to create a unique fingerprint for each image. For COSPIDS, the tip of the rostrum and spiracles were used as the three reference points. The user then marks the spot pattern of the organism, working to select the most prominent spots and patterns first. After the fingerprint has been created the software runs it against the database to produce a ranked list of images based on how well they match the new image. A number of limitations became apparent when trialling i3S. The process of selecting fifty distinguishable features to create a unique fingerprint for each image was a unique benefit to i3S, however it was time-consuming. Furthermore, the act of the user manually selecting the reference points opens the process up to bias. By limiting the number of distinguishable features, users have to choose between marking obvious spot patterns on certain parts of the body and marking spots from all over the body of the skate. Given that the software is looking at the entire body of the skate, then it is vital that the distinguishing features should be relatively evenly spaced. Furthermore, it is difficult to pick distinguishing features in individuals with spots of similar size, shape and pigment levels. Finally, the limit to the number of distinguishing features poses even more challenges by opening up the potential for false matches in individuals with hundreds of spots and those with few spots.

## HOTSPOTTER

HotSpotter was the final software tested. Developed by Rensselaer Polytechnic Institute, it is a free pattern recognition software which works by running a selected region of interest (ROI or chip) against a database. The process of entering an image and comparing it is a simple process consisting of 4 basic steps: import, select the ROI, correcting the orientations and running query. After running a query, the software produces a query results table which ranks the chips by their similarity scores, with a higher score reflecting a closer match. The main advantage of HotSpotter is that it is simple and straightforward to use, taking very little time to import and process an image. Furthermore, by drawing a rectangle around the skate the effect of sample bias is removed as the user does not have to select the spot pattern.

## ALTERATION TO SOFTWARE FUNCTION IN COSPIDS PROJECT

The primary objective for utilising a pattern-recognition software was to use it to match skate. However, during the software trials, it became apparent that the neither i3S nor HotSpotter could correctly rank a matching image in the top rank each time. It was decided that the software would now be used to perform a preliminary sift, reducing the number of images to compare from 330+ images (or in some cases 600+ images) down to 20 at a maximum.

## TRIAL RESULTS

Two trials were run on both i3S and HotSpotter to test which software suited the COSPIDS project best in terms of functionally and efficiency. The first trial consisted of 31 images, of 12 individuals, which were taken at the correct angle with limited glare. The results of this test showed conclusively that HotSpotter had a higher chance of correctly matching the submitted image to an image of the

| Rank | i3S (%) | HotSpotter (%) |
|---|---|---|
| Top 1 | 19.35 | 67.7 |
| Top 3 | 41.9 | 87.1 |
| Top 5 | 51.6 | 96.8 |
| Top 10 | 61.3 | 96.8 |
| Top 15 | 77.4 | 96.8 |
| Top 20 | 90.3 | 100 |
| Top 25 | 93.5 | 100 |
| Top 30 | 100 | 100 |

*Table 2. Results of the initial software test using 31 images of 12 individuals. Percentage matches are calculated by dividing the number of images with a match in that particular rank by the total number of images and then multiplying this number by 100.*

| Rank | i3S (%) | HotSpotter (%) |
|---|---|---|
| Top 1 | 71.8 | 44.7 |
| Top 3 | 77.7 | 60.2 |
| Top 5 | 79.6 | 66 |
| Top 10 | 81.6 | 73.8 |
| Top 15 | 84.5 | 85.4 |
| Top 20 | 89.3 | 88.4 |
| Top 25 | 91.3 | 89.3 |
| Top 30 | 92.2 | 90.3 |

*Table 1. Results of the initial software test using 103 images of 28 individuals. Percentage matches are calculated by the same method used in Table 1.*

same individual, with the top-ranking image matching the submitted image 67.7% of the time. Table 1 compares the percentile chance of a matching image being in the top *nth* category.

The second trial consisted of 103 images of 28 individuals. These images were taken prior to the development of the protocol detailing how to take images for the purpose of photo-identification. Many of the images are taken at oblique angles, with glare and shadows over the body of the skate. In some, the spots at the back of the skate are hard to see, resulting in the spots at the front of the body being primarily used as distinguishing features in i3S. Table 2 displays the results which show that i3S performed better for this sample of images. The decline in HotSpotter's capability of identifying photos is due to the wide variation in the angle of the image and the presence of glare and shadows. Following a discussion about the results of the software trials it was agreed that HotSpotter was the best suited software to the COSPIDS project given that moving forward the majority of new images would follow the photography protocol created by Scottish Natural Heritage (SNH) and SAMS.

## DEVELOPING PHOTO-ID PROTOCOL (COSPIDS P2)

After identifying HotSpotter as the software best suited to the COSPIDS project, the next step was to prepare a protocol for entering images in to the new software. COSPIDS P2 clearly describes how to import images into HotSpotter and how to prepare them for matching. In order for the software to perform optimally it is important that all chips are orientated in the same way, with the rostrum pointing towards the bottom centre of the chip. In addition to discussing how to import images there was also a discussion on how images should be selected to be imported into the database. Ultimately it was decided that the best image from each capture event would be entered into the database. The best image is classed as that with most of the skate's body in the frame and that with the least shadow and glare. As a result, there are multiple images for individuals which have been recaptured multiple times.

## DEVELOPING PHOTOGRAPH CATALOGUE

Given that there is the potential for multiple images of the same individual to be in the database at the same time it was important to develop nomenclature to name the images. A protocol, called COSPIDS P1, was created to detail how new images should be named and the importance of following the established nomenclature. Each file was named following the following structure: COSPIDS individual Di number, date (YYMMDD format), time (HHMM format), 8 figure PIT tag number if available, chip number and the name of the original file. PIT tagging in the MPA began in March 2016 resulting in most images taken from March 2016 onwards having an associated PIT tag number. Where possible these images were matched to this number. Images of the PIT tagged individuals were then matched against the 251 individuals which were identified by eye in early 2017 as part of a preliminary study.

## SKILLS DEVELOPED

The knowledge gained and the skills acquired during this internship complement and enhance the skill set which I have obtained and continue to develop throughout my undergraduate degree. Listed below is a brief summary of the skills and attributes developed during the internship;

Adaptability

- Various adaptations were made to the project during the internship including altering the role of the software.

Critical thinking pertaining to potential future applications and limitations of the COSPIDS project

- Proposing Wildbook as an interactive and engaging way to communicate with the angling community and also the general public.
- Proposing methodology for future processing of new images to minimise backload and work load.

Data processing and cataloguing

- Processing approximately 800+ images to select the images to be used in the database and linking these images to the PIT tag database

Developed understanding of multiple pattern-recognition software's

- Acquired during the software testing period

Interpersonal and communication skills

- Working closely with Dr Jane Dodd and Dr Steven Benjamins to keep them up to date on all developments and issues in the testing and cataloguing process.

Self-reliance

## SYNOPSIS

As a result of this internship, a catalogue of approximately 400 individuals has been created. Of these only 140 do not currently have an associated PIT tag number, due to the images been taking prior to March 2016 (when PIT tagging began). It is hoped that these images will be matched to PIT tags in the future when the individuals are recaptured. Two databases, one for males and one for females, have been established following the decision that it would speed up the identifying process and allow HotSpotter to run quicker. There is a group of 18 fish for which no gender has been assigned, either due to an obscured image where the fish could have claspers but they can't be seen in the image or the fish is small hence it is difficult to determine its sex. These fish have been assigned their own database (unknown gender) until such time as when their gender can be determined after recapture. Where the gender of the fish in a new image is known then it will be run against the male or female database depending on the gender (e.g. a female skate will be run against the female database) and the unknown gender database until such a time when all the fish in the unknown gender database have been assigned a gender. If the gender of the fish in the photo is unknown then it will run against all three databases.

At the time of completing the internship the male, female and unknown genders databases of the 251 fish identified in the preliminary study (fish photographed by Ronnie Campbell up to October 2016) are ready to be entered into Skate Spotter v1.0 in HotSpotter. These contain 330+ images in the female database and 80+ images in the male database, accounting for multiple recaptures of the 251 fish. Two hundred and fifty further images are currently being compared to the original 251 individuals. These images will be presented for comparison against Skate Spotter v1.0 before eventual addition into Skate Spotter v1.1.