Computer-assisted Recognition Of Dolphin Individuals Using Dorsal Fin Pigmentations

Andrew GILMAN*, Krista HUPMAN*, Karen A. STOCKIN*, Matthew D. M. PAWLEY*
*Institute of Natural and Mathematical Sciences, Massey University, Auckland, New Zealand
A.Gilman@massey.ac.nz

***Abstract*—Ecologists commonly use photo-identification of individual animals to monitor the behaviour, state and health of a population, since it is a cost-effective technique that eliminates the need to physically capture and tag animals. With dolphins, the nicks and notches of the dorsal fin are typically used as the unique identifying features for each individual; however New Zealand common dolphins are relatively unmarked, so most of the population cannot be identified. Here, we investigate how computer vision can be used to extract information from the pigmentation patterns that are typically seen on adult common dolphin dorsal fins. We develop features that are relatively robust to changes in the fin orientation and compare the classification rates of 779 photos of 169 different adult common dolphins. Using pigmentation-based features, we correctly classified individuals 75% of the time, with our top-5 estimates containing the correct dolphin in 86% of the cases.**

***Keywords*—Photo-id, Feature extraction, Image classification, Dolphins, Cetaceans, Pigmentation, Recognition**

I. INTRODUCTION

The identification of individual organisms is often of great importance to biologists — it is commonly used when exploring behaviour, group dynamics demography, home-range and migration patterns, as well as providing population estimates. Photo-identification (photo-id) makes use of unique, naturally occurring marks, eliminating the need to physically capture or tag the organism [1]. Many species have distinctive, easily identifiable marks, for example, cheetahs are commonly recognized through their pelage spots [2], whisker spot patterns for polar bears [3], pigmentation spots for whale sharks [4], scales of eastern water dragons [5], body patterns of jewelled geckos [6].

Photo-id technique is popular for identification of cetaceans, such as dolphins [7] because it is considered a non-invasive and cost-effective approach; but it requires the animal's fin to show some type of permanent damage around the edge, such as nicks, tears and notches (see Figure 1). These markings are then used to uniquely identify individuals from the photographs. Unfortunately, some species are less prone to acquiring such permanent markings; for example, New Zealand common dolphins (*Delphinus* spp.) have a relatively low dorsal fin nick-and-notch rate of 46.3% [8], which means a large proportion of the population cannot be identified.

Ongoing work [9] has established that 95.3% of adult common dolphins have pigmentation patterns on their dorsal fins, with most adults having patches of hypo-pigmentation

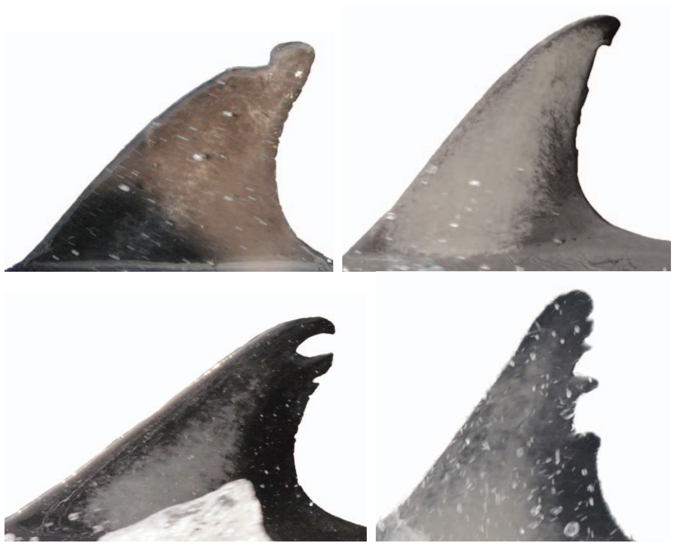

Fig. 1. Fins belonging to four different common dolphins with very distinctive nicks and notches that can be used to identify the individual.

(i.e. prevalent paler colours) on a relatively dark-coloured fin (see Figure 1 for a number of examples). This feature may be used to discriminate between individuals — an expert human observer was able to manually classify, with 88% accuracy, 670 images (taken between 2002 and 2013) of 186 different dolphins. However, manually identifying and sorting individuals and creating a catalogue is a time-consuming task. A computational method that could quantify the pigmentation patterns of the fins would enable machine learning techniques to (semi-)automate the recognition of individuals and relieve the researchers of a considerable burden. It would also allow the creation of catalogues that include a very large number of individuals. Without computer-assisted recognition this would be impractical, as matching new images to such a catalogue would take a significant amount of time.

The main purpose of this work is to investigate how computer vision can be used to improve the speed and robustness of the identification process based on information extracted from pigmentation patterns on dorsal fins.

II. PROBLEM OVERVIEW

Taking high quality photographs of common dolphins dorsal fins is difficult, even for experienced photographers. Common dolphins spend most of the time underwater and break the

978-1-5090-2748-4/16/$31.00 ©2016 IEEE

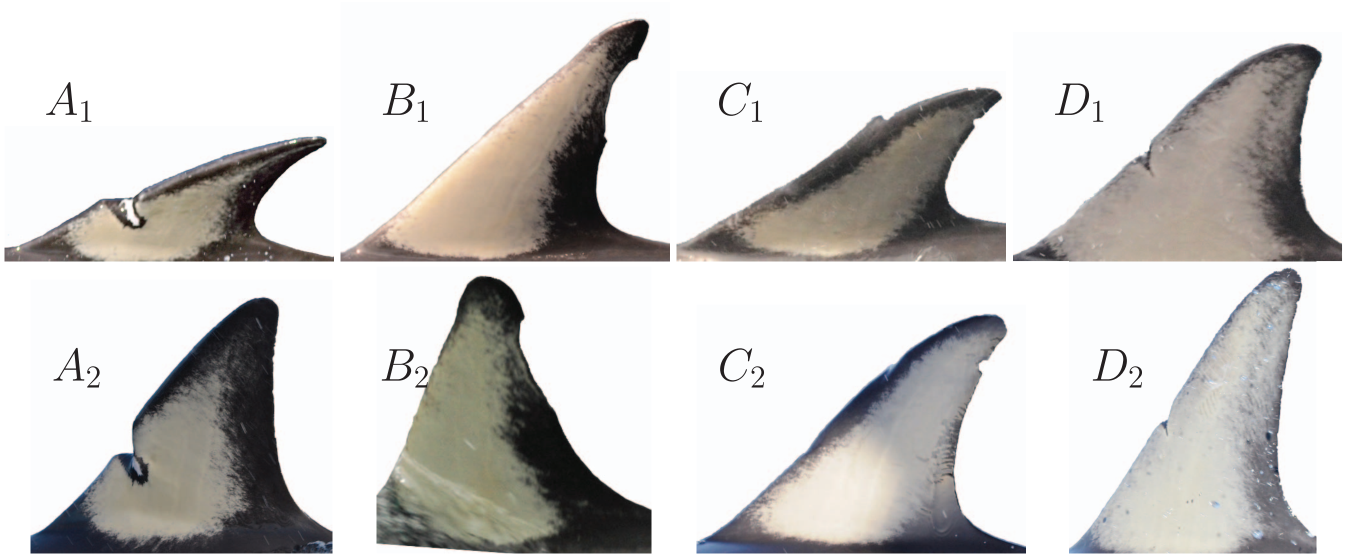

Fig. 2. Fins of four individual dolphins (one individual per column) captured from two different perspectives show how its apparent shape may be dramatically transformed. Also showing: A_1 and A_2 - exposure differences, C_2 - uneven illumination, B_2 - translucent occlusion, A_1 - minor specular highlights, A_1, D_1 and D_2 - minor droplets of water appearing as spots.

water surface for only brief periods (typically a fraction of a second). The photographer often has little idea of where the animal will surface in advance and has very short time to reposition the camera, adjust zoom, focus and take a photograph. The resulting images are seldom of perfect quality and suffer from motion and out-of-focus blurring, poor resolution (no time to zoom in on the fin) and also exposure problems, as lighting conditions can be quite variable. In addition, the photographs often suffer from occlusions of the fin by waves/sea spray formed as the animal breaks the water (see Figure 1) or water sheeting over the skin, forming a semi-translucent layer that is prone to showing reflections and specular highlights (see Figure 2 B_2). This video [10] shows the difficulty of capturing these animals on camera even under perfect conditions (sunny day, no swell and minimal wind).

Another problem inherent to the photographic process is perspective distortion. Ideally, researchers want to take photographs at an orthogonal angle to the fin; however, in practice the animal's angle relative to the photographer (rotation around the yaw axis) can vary greatly. In addition, photographs are frequently taken from a considerable height above the water, effectively introducing a rotation around the dolphin's roll axis. For close-up shots, these rotations introduce a considerable amount of perspective distortion (see Figure 2 for examples of extreme geometric deformations and non-uniform lighting).

Because of such a wide range in possible image quality of field photographs, researchers apply some strict grading protocols, for example, based on photographic-quality criteria with the aim of minimising bias and reducing misclassification [11], [8] when using prominent nicks and notches on the trailing (and some times leading) edge of the fin. Such features are relatively unaffected by blurring, occlusions (of other parts of the fin), poor exposure and illumination problems.This results in photographs that are sufficient to identify the individual via nicks and notches, but are of very poor quality otherwise, passing the grading.

Any pigmentation quantification method, for the purpose of individual identification, must be able to cope with such image quality problems. A range of techniques for computer-assisted photo identification using patterns have been proposed (I^3S [12], HotSpotter [13], [14], Wild-id [15], etc.); however, many of these rely on high-contrast patterns (spots, stripes, etc), which can be identified using standard computer vision features such as Scale-Invariant Feature Transform (SIFT) or Speeded Up Robust Features (SURF), or their variants. However, the pigmentation patterns on the dolphins lack sharp edges/spots (or high gradients in general), so such features are not really suitable, especially when considering the image quality problems.

The need, therefore, is to find features that can work with variable quality images and do not require easily identifiable, high-contrast, high-gradient salient points. Here, we investigate an approach where we extract some fairly 'coarse' features and perform multi-class classification.

III. FEATURE EXTRACTION

We hypothesised that subdividing the fin and summarising pixel values within each division with a robust statistic would result in feature values that were fairly invariant to blurring, poor resolution, occlusions, specular highlights and reflections. We attempted to find a subdivision scheme that would give a good trade-off between discriminatory power and robustness to these image quality problems.

However, selection of the subdivision scheme was complicated by the presence of perspective distortion. Corresponding segments should encompass pixels that belong to the same area of the fin for a wide variety of geometric transformations

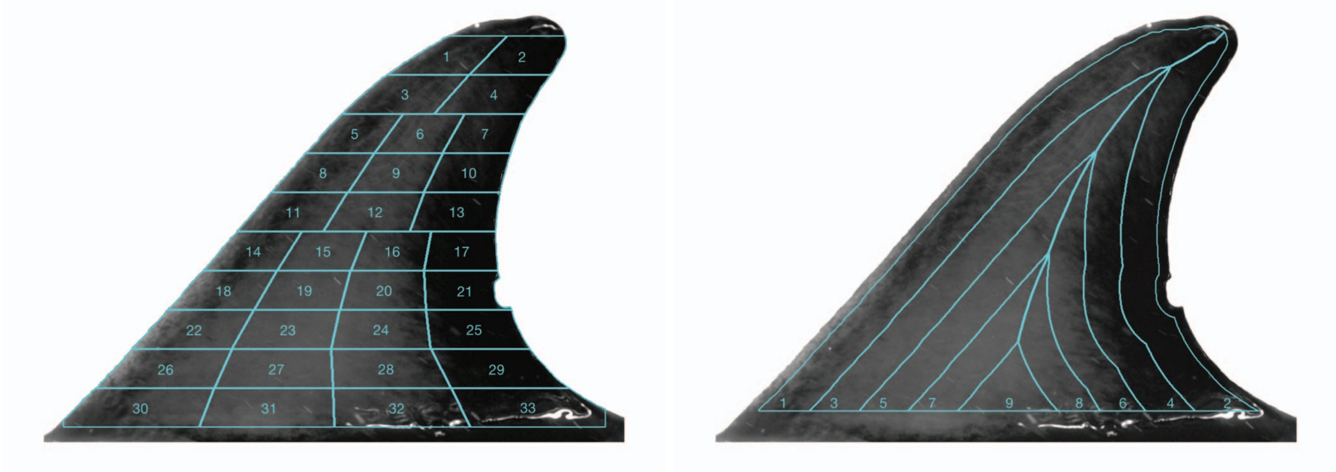

Fig. 3. Grid and contour based subdivision schemes.

encountered in practice. One way to accomplish this would be to transform the fins (warp the images) to some standard pose. Even though dorsal fins lack salient features that can be robustly identified, we have shown in the past [16] that it is possible to register two fins without these, by performing registration on the fin contours using the Iterative Closest Point algorithm. Finding a standard pose is almost impossible, however, as the fins can vary greatly in morphology. Since the fins can vary greatly in morphology, the only part of the fin that can be robustly standardized across images is the base of the fin. So the subdivision scheme must be relatively robust to the remaining distortion (after rotation of the base has been corrected for).

A. Robust subdivisions of the fin

We explored a number of different ideas and settled on two subdivision schemes. The first does not require finding any salient features and derives the subdivision from the fin outline (a contour-based division); the second utilises the fact that the base of the fin can be (approximately) identified from the point where the leading edge of the fin flows into the animal's body, forming an inflection point in the outline (a grid-based division).

1) Contour-based segments based on the closest distance from the fin edge: A distance transform was computed on the outline of the fin — each pixel that fell within the fin area was allocated a value corresponding to the Euclidean distance to the nearest point on the fin outline. Equidistant iso-contours were computed and used to subdivide the contour into 5 equal-width segments running along the fin outline and located progressively further towards the centre (see right half of Figure 3). These divisions were made in the hope that the area between these iso-contours would be robust to moderate changes in dorsal fin orientation.

To give these features more discriminating power, the fin's medial axis was computed by tracking the ridge of the distance transform (essentially a set of points equidistant from fin's leading and trailing edges). This medial axis was used to further subdivide the 4 of the 5 segments (except the centre-most segment) into left and right halves, resulting in 9 contour-based image patches.

2) Grid segments: As the base of the dorsal fin can be identified in a fairly robust manner, the fin was divided into 10 equal segments along an axis perpendicular to the base, with the upper limit of the segments was defined as half the segment's height from the top of fin (we ignored top-most part of the fin as it often contained specular highlights). The bottom 5 sections were further subdivided into 4 patches, the 3 middle horizontal segments were subdivided into 3 patches, and the top 2 horizontal segments were split in half (see left half of Figure 3), resulting in 33 image patches arranged in a grid-like fashion.

B. Pixel value normalisation

All images were converted to greyscale and their pixel intensity values were normalised by subtracting the mean and dividing by the standard deviation of all pixels that fell within the fin area to account for differences in camera's exposure settings and global illumination changes.

C. Summary statistics

The following summary statistics were calculated on the distribution of normalised greyscale pixel intensities within each of the 42 patches: mean, median and interquartile range (IQR). In addition, the following inter-divisional features were calculated:

std(gridMeans)	std(gridMedians)	std(gridIQRs)
IQR(gridMeans)	IQR(gridMedians)	IQR(gridIQRs)
		mean(gridIQR)
		median(gridIQR)
std(contMeans)	std(contMedians)	std(contIQRs)
IQR(contMean)	IQR(contMedians)	IQR(contIQRs)
		mean(contIQRs)
		median(contIQRs)

Where std(gridMeans) denotes standard deviation of the distribution of the 33 grid segment means, std(contMeans) denotes standard deviation of the distribution of the 9 contour segment means, etc. In total, 142 features were initially used to characterise each image.

IV. Evaluation of subdivision schemes

Ideally, the fin shape should be subdivided into a set of segments, where each segment will encompass the same physical area of the fin, even if the image is captured from different perspectives (within some practical range). Here we perform a simple experiment to demonstrate the effect of perspective distortion on the proposed subdivision schemes.

Two images of fins of the same individual are registered to compute the projective transformation required to warp one image (moving) into the coordinate system of the other (reference). Then our subdivision scheme is applied to the original moving image and transformed using the above transformation into the reference coordinate system. Now we can apply the subdivision scheme to the reference image and compare the two. The better these two match, the more invariant the subdivision scheme is to the transformation.

Figure 4 shows the result of performing this experiment for two individuals with slightly different fin morphology. The first one (top row) exhibits a moderate amount of distortion. We can also make another interesting observation, stemming from poor cropping of the moving image. The inflection point where the leading edge of the fin flows into the body has been cropped out, resulting in a small error in location of the base of the fin. We can observe that this error causes a small mismatch for the grid segments (bottom left of the fin), however this error does not have much effect on the contour segments (except they end prematurely at the bottom as that is where the moving image has been cropped).

Second row of Figure 4 shows another individual; however, here we have a much greater perspective distortion. This is not a typical case, but such extreme transformations do occur. We can observe that the grid subdivision still has very good overlap, showing that it is robust to the most extreme transformations encountered. Contour based subdivision, on the other hand, has somewhat failed. The segments closest to the fin outline are more robust than the segments towards the fin centre. The medial axis, computed from the local maximums of the distance transform is actually the least robust point — it has moved across by the largest distance. This leads us to believe that the grid segments should be more robust to extreme perspective distortion than the contour segments and should perform better in classification experiments.

V. Evaluation of classification

A. Dataset

We evaluate the effectiveness of the proposed features in practice by recognising individuals in a catalogue of photographs of common dolphins (collected within the Hauraki Gulf) using linear discriminant analysis (LDA). The catalogue contains animals with distinctive nicks and notches that allowed them to be identified with high certainty, providing us with ground truth for our experiments of identification using pigmentation alone. Dolphin photographs within the catalogue were taken from distances of 2 to 25 meters, using Nikon D90 and D7000 digital SLR cameras, equipped with Nikon 100-300 mm and 100-400 mm lenses, respectively. Photo-id sessions were performed using two separate vessels with platform heights of 0.5 and 2 meters respectively. Only one side of the dorsal fin (left) was photographed as some individuals exhibited only minor nicks and notches, which were not considered to be detectable from both sides.

The dataset contains 779 photographs of 169 different adult common dolphins, acquired over a number of years (2010-2013) during 187 photo-id sessions. Figure 5 shows the distribution of number of images per individual.

B. Feature ranking

Only a subset of the full set of features were chosen to represent the fin. The subset was selected by initially ranking the importance of each feature within LDA using correlation adjusted t-scores (cat scores) [17]. When there is no correlation between features, the standardized distance (t-score) between class centroids and the group centroid is considered to be a natural criterion to determine which features should be selected for a classifier [18] – the cat-score is an extension that generalizes the t-score to account for correlation between features and is therefore suitable to rank feature importance for linear discriminant analysis (LDA). Because of the high dimensionality of the feature dataset and relatively small number of images per dolphin, a 'shrinkage' method was used to stabilize the estimate of the covariance matrix. We used a regularized linear discriminant analysis procedure with a James-Stein shrinkage estimate of the covariance matrix implemented in the R package *sda* [19]. We shrunk the off-diagonal elements of the covariance matrices towards zero in an effort to obtain a biased but less variable estimator.

C. Training

Linear Discriminant Analysis was used to classify the images. The LDA classification rate was determined by leave-one-out cross validation (LOOCV). After taking one instance out, majority of classes had 3 or less members in the training set and 32 had only 2 members.

The LOOCV classification success was calculated for the percentage of times: (i) the top choice from the LDA was correct (top-1), and (ii) the top-5 choices contained the correct individual (top-5). The number of features used in the LDA was considered a hyperparameter and was optimized based on the top-1 LOOCV results.

D. Results & Discussion

Table I shows the results of leave-one-out cross validation of the LDA classification rate for each of the proposed subdivision scheme, as well as combining the features from

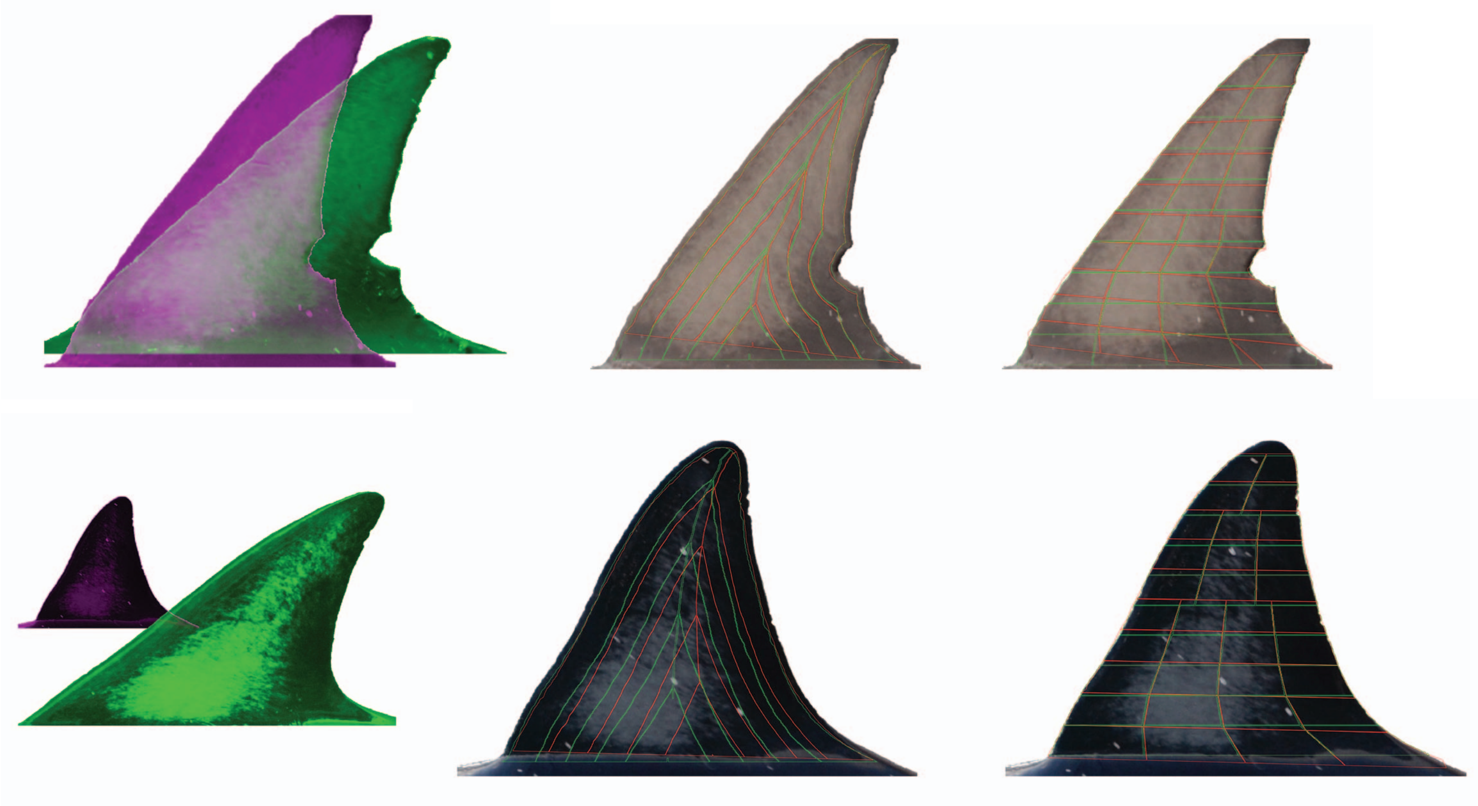

Fig. 4. The robustness of the patch positions in different images of an individual (one dolphin per row is shown). The 1^{st} column shows the original image shapes (green fin is registered against the red). Features are applied to the unregistered images. The 2^{nd} and 3^{rd} columns show the position of the patches when transformed onto the reference image.

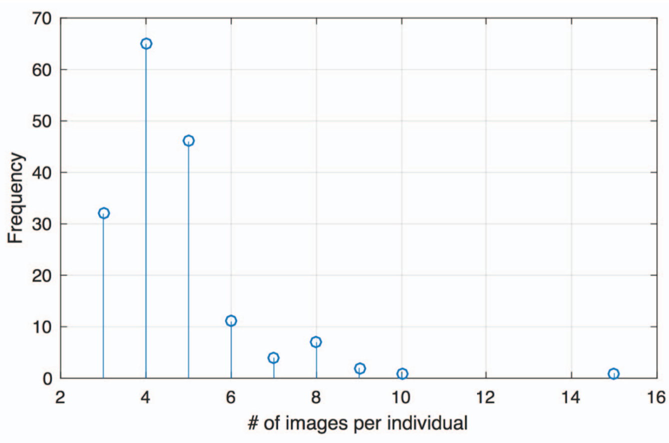

Fig. 5. Frequency of individuals appearing in the dataset. Mean = 4.6.

both schemes. The results with and without the feature selection procedure (described in section V-B) are provided for comparison.

The classification results suggest that the grid-based subdivision features (71.2% top-1 and 83.7% top-5 correct classifications) provide greater discriminatory power than the contour-based subdivisions (53.5% and 77.3%) on average over wide range of fin morphologies and perspective distortions. Using

TABLE I
RESULTS OF LOOCV AS A PERCENTAGE OF CORRECT CLASSIFICATIONS FOR EACH SUBDIVISION SCHEME.

Subdivision type	# of features used / total	% correct top-1	% correct top-5
Contour-based	35/35	51.5	77.2
Contour-based	27/35	53.5	77.3
Grid-based	107/107	69.2	81.0
Grid-based	76/107	71.2	83.7
Grid + Contour	142/142	73.0	84.5
Grid + Contour	97/142	**75.5**	**86.3**

the features from both subdivision schemes together offers a small improvement on using grid segments alone (71.2% $\rightarrow$ 75.5% for top-1 and 83.7% $\rightarrow$ 86.3% for top-5). In all cases, reducing the number of features, using the feature selection processes described earlier, marginally improved classification rates (by around 2%), and the best classification rates were achieved after reduction in feature number by between 20 and 30 per cent.

At first glance, 75% may not seem like a particularly convincing classification rate, but one should keep in mind that there were 169 different classes to choose from, with most individuals having 4 or fewer images. Indeed, an experiment where a human expert tried to classify individuals within the dolphin catalogue using pigmentation patterns alone (nicks and notches were artificially removed in the images using image in-painting to blind the participant) had a success-rate of around 88% [9].

VI. CONCLUSIONS

The classification of common dolphins using pigmentation patterns from field photographs is a difficult task due to the inherently poor quality of the images. However, it would greatly benefit ecologists working in this area: ecologists estimate that there are thousands of different common dolphin individuals residing within the Hauraki Gulf, yet the sparsity of individuals with nicks and notches has restricted the catalogue to only 186 individuals.

This work has shown that there is clearly pigment signal available — the model that used Grid + Contours had a top-1 classification rate of 75.5%, this is approximately 113 times higher than classifying individuals at random (i.e. a lift of 113) and a top-5 classification rate of 86.3%, which nears the 88% success rate achieved by a human expert. From a functional perspective, the ability to quickly classify most new images by examining a subset of the 5–10 most likely individuals will save the ecologist considerable time.

We are aiming to improve the classification success rate even further in the future by combining pigmentation features with fin shape metrics, as well as features based on nicks and notches if they are present.

REFERENCES

[1] Bearzi, G., Politi, E., Agazzi, S., Bruno, S., Costa, M., and Bonizzoni, S., "Occurrence and present status of coastal dolphins (*Delphinus delphis* and *Tursiops truncatus*) in the eastern Ionian Sea", *Aquatic Conservation: Marine and Freshwater Ecosystems*, vol. 15(3), pp. 243–257, 2005.

[2] Kelly, M. J., "Computer-aided photograph matching in studies using individual identification: an example from Serengeti cheetahs", *Journal of Mammalogy*, vol. 82(2), pp. 440–449, 2001.

[3] Anderson, C. J., Lobo, N. D. V., Roth, J. D., and Waterman, J. M., "Computer-aided photo-identification system with an application to polar bears based on whisker spot patterns", *Journal of Mammalogy*, vol. 91(6), pp. 1350–1359, 2010.

[4] Brooks, K., Rowat, D., Pierce, S. J., Jouannet, D., and Vely, M., "Seeing spots: photo-identification as a regional tool for whale shark identification", *Western Indian Ocean Journal of Marine Science*, vol. 9(2), pp. 185–194, 2010.

[5] Gardiner, R.Z., Doran, E., Strickland, K., Carpenter-Bundhoo, L. and Frre, C., "A face in the crowd: a non-invasive and cost effective photo-identification methodology to understand the fine scale movement of eastern water dragons", *PloS one*, vol. 9(5): e96992, 2014.

[6] Knox, C.D., Cree, A. and Seddon, P.J., "Accurate identification of individual geckos (*Naultinus gemmeus*) through dorsal pattern differentiation", *New Zealand Journal of Ecology*, vol. 37(1), pp. 60-66, 2013.

[7] Würsig B. and Jefferson T. A., Methods of photo-identification for small cetaceans, no. Special Issue 12. 1990.

[8] Hupman K., Photo-identification and its application to gregarious dolphins: Common dolphins (*Delphinus sp.*) in the Hauraki Gulf, New Zealand. PhD thesis, Massey University, Auckland, New Zealand, 2016.

[9] Pawley, M.D.M., Hupman, K.E., Stockin K.A. and Gilman, A., "Dorsal fin pigmentation patterns can be used as a sole feature for photo-identification of common dolphins (*Delphinus* sp.)", unpublished (*in prep*).

[10] Coastal-Marine Research Group, Massey University "A day with the dolphins", https://vimeo.com/183759068

[11] Tyne, J.A., Pollock, K.H., Johnston, D.W. and Bejder, L., "Abundance and survival rates of the Hawaii Island associated spinner dolphin (*Stenella longirostris*) stock", *PloS one*, vol. 9(1): e86132, 2014.

[12] I3S Pattern v4.0 User Manual, available from http://www.reijns.com/i3s/download/I3S_download.html, *accessed on 20 September 2016*.

[13] HotSpotter software, available from http://cs.rpi.edu/hotspotter/

[14] Crall, J.P., Stewart, C.V., Berger-Wolf, T.Y., Rubenstein, D.I., Sundaresan, S.R., "HotSpotter — Patterned species instance recognition", in proceedings of 2013 IEEE Workshop on Applications of Computer Vision (WACV), Tampa, FL, 2013, pp. 230-237.

[15] Bolger D., Vance B., Morrison T., and Farid H., "Wild-id user guide: Pattern extraction and matching software for computer-assisted photographic mark-recapture analysis", available from http://software.dartmouth.edu/Macintosh/ Academic/Wild-ID_1.0.0.zip.

[16] Gilman, A., Dong, T., Hupman, K., Stockin, K. and Pawley, M., "Dolphin fin pose correction using ICP in application to photo-identification", in proceedings of 28th International Conference on Image and Vision Computing New Zealand (IVCNZ '13), Wellington, New Zealand, November 2013, pp. 388-393.

[17] Ahdesmäki, M., and Strimmer K., "Feature selection in omics prediction problems using cat scores and false nondiscovery rate control", *The Annals of Applied Statistics*, vol. 4(1), pp. 503-519, 2010.

[18] Fan, J., and Fan, Y., "High dimensional classification using features annealed independence rules", *Annals of statistics*, vol. 36(6), pp. 2605, 2008.

[19] Ahdesmäki, M., Zuber, V., Gibb, S. and Korbinian S., "sda: Shrinkage Discriminant Analysis and CAT Score Variable Selection", R package version 1.3.7., available from http://CRAN.R-project.org/package=sda