

Automated Identification of Individual Great White Sharks from Unrestricted Fin Imagery

Benjamin Hughes
benjamin.hughes@bristol.ac.uk

Tilo Burghardt
tilo@cs.bris.ac.uk

Department of Computer Science
University of Bristol
Bristol, UK

Abstract

The objective of this paper is automatically to identify individual great white sharks in a database of thousands of unconstrained fin images. The approach put forward appreciates shark fins in natural imagery as smooth, flexible and partially occluded objects with an individuality encoding trailing edge. In order to recover animal identities therefrom we first introduce an open contour stroke model which extends multi-scale region segmentation to achieve robust fin detection. Secondly, we show that combinatorial spectral fingerprinting can successfully encode individuality in fin boundaries. We combine both approaches in a fine-grained multi-instance recognition framework. We provide an evaluation of the system components and report their performance and properties.

1 Introduction

Recognising individuals repeatedly over time is a basic requirement for field-based ecology and related life sciences [25]. In scenarios where photographic capture is feasible and animals are visually unique, biometric computer vision offers a non-invasive identification paradigm for handling this problem class efficiently [23]. The principal aim of such systems is the recovery of animal identities from images that reveal unique aspects of visual animal appearance. To act as an effective aid to biologists, systems are required to operate reliably on unconstrained, natural imagery as to facilitate adoption over widely available, manual or semi-manual identification software [22, 24, 61, 62, 67].

In this paper we propose a visual identification approach for great white shark fins as schematically outlined in Figure 1, one that is applicable to unconstrained fin imagery and fully automates the pipeline from image feature extraction to matching of identities. We note that fin shape information has been used in the past to track individual great white sharks over prolonged periods of time [10] or global space [9]. Recently fin re-identification has also been conducted semi-automatically [12, 63]. However, to the best of our knowledge the proposed system is the first fully automated contour-based animal biometrics system.

We pose the associated vision task as a fine-grained, multi-instance classification problem for flexible, smooth and partly occluded object parts. ‘Fine-grained’ in that each individual fin, described by a characteristic jagged trailing edge, is a subclass of the parent class great white shark fin. ‘Multi-instance’ since the system should be able to assign potentially multiple semantic labels to an image, each label corresponding to an individual

shark present. ‘Flexible’ and ‘smooth’, since fins may bend and lack distinctive colour or 2D texture. Whilst some sharks carry fin pigmentation, not all do and its permanence is disputed [60]. Finally, fin detection poses a part recognition problem which region-based recognition on its own fails to tackle (see Figure 2): species-typical shaped fins are visually smoothly connected to the shark body whilst being partly occluded by the water line and white splash. Figure 1 shows examples of the dataset and outlines our solution pipeline – from image to individual shark ID.

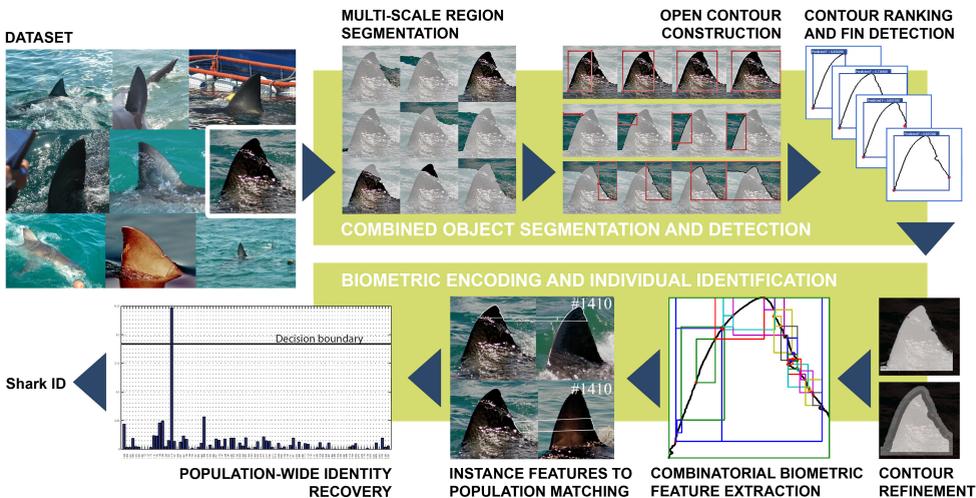


Figure 1: **SYSTEM OVERVIEW:** We perform a coarse and a fine-grained recognition task. The first is to simultaneously segment and detect shark fins, and the second is to recognise individuals. We begin by segmenting an image into an ultrametric contour map, before partitioning boundaries into sets of open contours. We then train a random forest to rank contours and detect fin candidates based on normal information and opponentSIFT features. This forms the basis for computing individually distinctive contour features, which are matched against a population database to recover shark identities.

2 Related Work and Rationale

Smooth Object Recognition. Smooth object recognition traditionally builds on utilising boundary and internal contour features, and configurations thereof. Recent approaches [2, 3] extend these base features by mechanisms for regionalising or globalising information, and infer object presence from learning configuration or grouping classifiers. A prominent, recent example is Arandjelovic and Zisserman’s ‘Bag of Boundaries (BoB)’ approach [3] which employs multiscale, semi-local shape-based boundary descriptors to regionalise BoB features and subsequently predict object presence. A related, more efficient boundary representation is proposed in [4], which focusses on a 1D semi-local description of boundary neighbourhoods around salient scale-space curvature maxima. This description is based on a vector of boundary normals (Bag of Normals; BoN). However, experiments in [4] are run on images taken under controlled conditions [18], whilst in our work, in common with [3], we have the goal of separating objects in natural images and against cluttered backgrounds (see Figure 1).

Fin Segmentation Considerations. The biometric problem at hand requires an explicit, pixel-accurate encoding of the fin boundary and sections thereof to readily derive individually characteristic descriptors. To achieve such segmentation one could utilise various approaches, including 1) a bottom-up grouping process from which to generate object hypotheses for subsequent detection [11, 19, 24, 54], or 2) a top-down sliding window detector such as [14, 17, 68] and then segment further detail, or 3) combining the two simultaneously [9]. We select the first option here since boundary encoding is intrinsic, and bottom-up, efficient and accurate object segmentation has recently become feasible. Arbelaéz et al. in [9] introduce a fast normalised cuts algorithm which is used to globalise local edge responses produced by the (fast) structured edge detector of Dollár et al. [15]. However, since fins represent open contour structures we require some form of (multi-scale) open contour generation which is proposed, similar to [9], by stipulating keypoints along the closed contours of the ultrametric map as generated by [9]. Our proposed contour stroke model (see Section 3) combines 1D shape information of open contour sections and 2D region information around them to segment the target fin structures. Note that these are objects which are not present as segments at *any* level of the underlying ultrametric contour map, as Figure 2 illustrates.



Figure 2: FIN DETECTION AS OPEN CONTOUR STROKES: Multi-scale 2D region-based segmentation algorithms [9] on their own (left images) regularly fail to detect the extent of fins due to visual ambiguities produced by shark body, water reflections or white splash. Thus, often no level of the underlying ultrametric contour map captures fin regions. We suggest combining properties of the 1D (open) contour segment shape with local 2D region structure in a contour stroke model to recognise the fin section (shown in solid white).

Biometrics Context. Most closely related within the animal biometrics literature are the computer-assisted fin recognition systems; DARWIN [32] and Finscan [27]. DARWIN has

been applied to great white sharks [10, 33] and bottlenose dolphins [36] while Finscan has been applied to false killer whales [6], bottlenose dolphins [7] and great white sharks, among other species [10]. However both differ significantly from our work in that they rely on user interaction to detect and extract fin instances, while fin descriptions are sensitive to partial occlusions since they are represented by single, global reference encodings. Additionally, in the case of DARWIN, fin shape is encoded as 2D Cartesian coordinates, requiring the use of pairwise correspondence matching. By contrast, we introduce an occlusion robust vector representation of semi-local fin shape (see Section 4). As in [33], this allows multiple images of individuals to be held in tree-based search structures of the population, which allow identity discovery in sub-linear time.

Contribution and Paper Structure. The paper has three key technical contributions, that is 1) a contour stroke model trained for fin detection combining a partitioning of ultrametric contour maps with normal descriptors and dense local features, 2) a biometric encoding scheme for contour sections, and 3) a real-world application to individually identify great white shark fins evaluated on a dataset of thousands of images.

The remainder of the paper will detail the methodology and algorithms proposed, it will then report on application results and briefly discuss our approach in its wider context.

3 Contour Stroke Object Model

In this section we describe our contour stroke model for bottom-up fin detection. It constructs fin candidates as subsections (or ‘strokes’) of contours in partitioned ultrametric maps and validates them by regression of associated stroke properties. The approach progresses in three stages: 1) image segmentation is used to detect and group object boundaries at multiple scales into an ultrametric contour map, 2) salient boundary locations are detected and used to partition region boundaries into contour sections called strokes, 3) strokes are classified into fin and background classes based on shape, encoded by normals, and local appearance encoded by opponentSIFT features. Figure 3 illustrates the fin detection approach in detail.

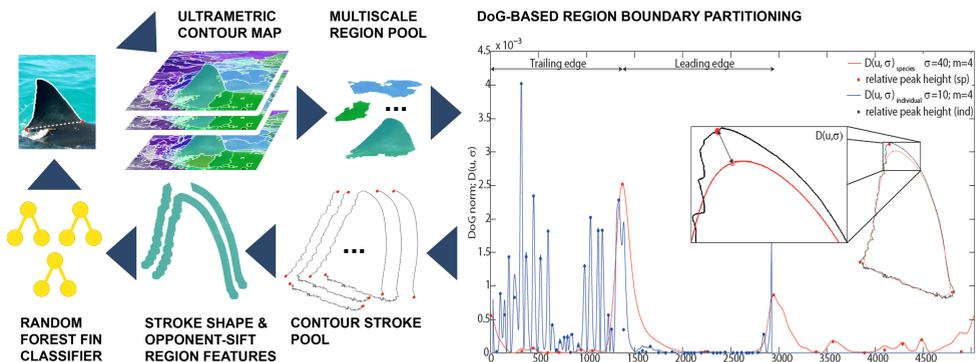


Figure 3: FIN DETECTION MODEL: Partitioning the (closed) 2D region structures from across *all* levels of the ultrametric contour map via DoG-generated keypoints (rightmost visualisation) yields a pool of 1D (open) contour strokes, whose normal-encoded shape and nearby opponentSIFT descriptors feed into a random forest regressor to detect fin objects.

1. Image Segmentation. We use the method and code of [5] to generate a region hierarchy in the form of an ultrametric contour map. This provides sets of closed regions for any chosen level-threshold in the range $[0, 1]$. Starting with the whole image, we descend the hierarchy to a pool of 200 unique regions. Similar to [10], we then employ fast region rejection to remove areas too small to represent a fin, or too similar to another region¹. We then rank the remaining regions, again by their location in the hierarchy, and retain the top k regions, choosing $k = 12$ empirically for the results produced in this paper.

2. Generating Fin Candidates. In almost all cases, image segmentation produces a single region, within a set, that provides a high recall description of the fin’s external boundary. However, in cases where the boundary between the fin and the body is visually smooth, segmentation tends to group both in a single region. The global appearance of such regions can vary dramatically, making 2D structures unsuitable targets for recognition. By contrast, locations along the 1D contour of regions provide discontinuities in curvature suitable for region sub-sectioning and stroke generation. We detect boundary keypoints using the Difference of Gaussian (DoG) corner detector of Zhang et al. [49]. Letting $C(u) = (x(u), y(u))$ represent a planar curve, the corner response function is given by the evolution difference of two Gaussian smoothed planar curves, measured using the distance $D(u, \sigma)$:

$$\begin{aligned} D(u, \sigma) &= [DoG * x(u)]^2 + [DoG * y(u)]^2 \\ &= [G(u, m\sigma) * x(u) - G(u, \sigma) * x(u)]^2 + [G(u, m\sigma) * y(u) - G(u, \sigma) * y(u)]^2 \end{aligned} \quad (1)$$

where $G(u, \sigma)$ is a zero mean Gaussian function with standard deviation σ , and $m > 0$ is a multiplication factor. Viewed as a bandpass filter, by varying m and σ , the operator can be tuned to different frequency components of contour shape. For keypoint detection (visualised rightmost in Figure 3), we resample contours to 128 pixels and compute D using $\sigma = 1$ and $m = 4$ before ranking the local maxima of D by their prominence. This suppresses locally non-maximal corner responses, before selecting the n largest as boundary keypoints. Choosing small values of σ ensures accurate keypoint localisation while a relatively large value of m ensures that the n largest maxima of D correspond to globally salient locations. We then generate fin candidates as contour strokes by sampling the region contour between every permutation of keypoint pairs.

This results in a pool of $N_c = (n^2 - n)k$ strokes per image. We set n by assessing the *achievable quality* (the quality of the best candidate as selected by an oracle) of the candidate pool with respect to the number of candidates. We denote this fin-like quality of stroke candidates by F_{inst} . Evaluated with respect to a human ground truth contour, we use the standard F -measure for evaluating contour detections based on bipartite matching of boundary pixels [26]. We observe that average achievable quality does not increase beyond $n = 7$ given the described DoG parameterisation and therefore use this value to define N_c . The result is that, on average, we obtain 504 candidates per image, with an average achievable quality of $F_{\text{inst}} = 0.97$ measured against human labelled ground truth contours for 240 randomly selected images. By means of comparison, the average achievable quality of the pool of $k = 12$ closed region contours is $F_{\text{inst}} = 0.75$.

3. Evaluation of Candidate Classification. For training and testing the candidate classifier,

¹Any region with a boundary length of less than 70 pixels is discarded, before the remainder are clustered into groups where all regions in a cluster have an overlap of 0.95 or more. Within each cluster, we rank regions according to the level in the hierarchy at which they first appeared, retaining the top ranked region in each cluster.

240 high visibility images (where the whole fin could clearly be seen *above* the waterline) are selected at random and then randomly assigned to either a training or validation set, each containing 120 images. Each image is resized to an area of 0.05 megapixels and ground truth fin boundary locations are labelled by hand using a single, continuous contour, 1 pixel in width. Each contour section is described by a 160-dimensional feature vector consisting of two components, contributing 2D and 1D distinctive information, respectively. The first is a bag of opponentSIFT [65] visual words (dictionary size 20) computed at multiple scales (patch sizes 16, 24, 32, 40) centred at every pixel within a distance of 4 pixels of the contour section, and is used to describe the local appearance of fin contours. The second describes contour shape using a histogram of boundary normals consisting of 20 spatial bins and 8 orientation bins. The two components are L_2 normalised and concatenated to produce the final descriptor. A random forest regressor [44] is then trained to predict the quality of fin hypotheses where the quality of individual candidates is assessed using the F -measure as computed using the BSDS contour detection evaluation framework [26]. Following non-maximum suppression with a contour overlap threshold of 0.2, a final classification is made by thresholding the predicted quality score.

Given an image, the detector produces a set of candidate detections, each with a predicted quality score F_{inst} . We accept a candidate if its F_{inst} score exceeds a threshold t . We use average precision (AP) as an evaluation metric for detection performance. Results for fin-contour detection are shown in Table 1. For almost all fin instances (up to 98%), a high quality candidate is generated and recognised as a correct fin. This result demonstrates the effectiveness of our contour stroke model for the task at hand.

Quality (F_{inst})	>0.7	>0.8	>0.85	>0.9
AP	0.98	0.98	0.95	0.85

Table 1: FIN DETECTION RESULTS

Edge Refinement. Our segmentation and contour partitioning framework produces descriptions of fin contours, but it does not resolve *to sufficient resolution* the fin shape along trailing edge and tip vital to distinguishing individuals within shark populations [4, 9]. To recover this detailing we apply border matting in a narrow strip either side of region boundaries using the local learning method and code of Zheng et al. [40]. This produces an opacity mask α which defines a soft segmentation of the image ($\alpha_i \in [0, 1]$). We obtain a binary assignment of pixels (by threshold 0.5) to separate fin and background, and extract the resulting high resolution contour of best Chamfer distance fit^2 as a precursor to biometric encoding.

4 Biometric Contour Encoding

In this section we develop a method of encoding smooth object shape suited to individual white shark fin representation. It enables efficient and accurate individual recognition while being robust to noisy, partially occluded input generated by automatic shape extraction.

Global shape descriptions, as used in [62], maximise inter-class variance but are sensitive to partial occlusions and object-contour detection errors, while the removal of nuisance

²We find the two locations on the boundary of the largest connected foreground component that have the shortest Euclidean distance to the two ends of the unrefined contour section. We partition the foreground region boundary into two sections using these locations and then measure each refined sections' similarity to the unrefined contour using Chamfer distance as a metric. The most similar section is taken as the final detected candidate.

variables such as in- and out-of-plane rotation rely upon computing point correspondences and inefficient pairwise image matching.

By contrast, the semi-local descriptions of [14, 15] are robust and allow efficient matching, but their encoding of inter-class variance will always be sub-maximal. Thus, we utilise *both* semi-local and global shape descriptions with a framework extending that used to generate fin candidates.

Generating Boundary Subsections. As a first step, we detect salient boundary keypoints on the extracted contour strokes to produce repeatably recognisable contour subsections that serve as descriptors. For keypoint detection we resample fin candidates to a fixed resolution of 1024 pixels and then compute $D(u, \sigma)$ in Equation 1 parameterised with $\sigma = 2$ and $m = 8$. Subdivision by these keypoints yields $\binom{50}{2} = 1225$ contour subsections³.

Boundary Descriptor. Following the generation of boundary subsections, the task is to encode their shape information. For this task we use the DoG norm (DoG_N) defined in Equation 1. It is well suited to the task of biometric contour encoding: first, the associated metric is suitable for establishing similarity between distance vectors D , meaning contour sections can be matched efficiently. Secondly, by varying the parameters σ and m , the description can be tuned to encode different components of the shape frequency spectrum. Third, the descriptor is rotation invariant and, as suggested by the correct matches shown in Figure 5, is robust to substantial changes in viewpoint.

Individual LNBNN Classification. Unlike in conventional object and image classification tasks, the availability of high quality, labelled training data for individual animals is often scarce. Thus, individuality information needs to be maintained throughout processing. As noted by Boiman et al. [16], information is lost in processes such as vector quantisation. For this reason, we utilise a scoring mechanism inspired by the local naive Bayes nearest neighbour (LNBNN) classification algorithm [17], and similar to that employed by [13] in the context of patterned species individual identification, to provide a recognition baseline.

Specifically, denoting the set of descriptors for a query object D_Q , for each query descriptor $d_i \in D_Q$, we find the two nearest neighbours ($NN_C(d_i), NN_{\bar{C}}(d_i)$) where C is the class of the nearest neighbour and \bar{C} is the set of all other classes. Query objects are then classified according to:

$$\hat{C} = \arg \max_C \sum_{i=1}^{|D_Q|} \left| |d_i - NN_{\bar{C}}(d_i)|^2 - |d_i - NN_C(d_i)|^2 \right| \quad (2)$$

This decision rule can then be extended to a multi-scale case. Letting $S = \{\sigma_1, \dots, \sigma_j, \dots, \sigma_v\}$ denote the set of scales for which we compute descriptors, the multi-scale decision rule linearly combines the contribution of the descriptors at each scale:

$$\hat{C} = \arg \max_C \sum_{j=1}^v w_j \cdot \sum_{i=1}^{|D_Q^j|} \left| |d_i^j - NN_{\bar{C}}(d_i^j)|^2 - |d_i^j - NN_C(d_i^j)|^2 \right| \quad (3)$$

Implementation Details. To achieve scale normalisation, each contour subsection is resampled to a fixed length of 256 pixels. DoG_N descriptors are then computed at (spectral) scales

³Taking as keypoints the $n = 48 + 2$ largest local maxima of D , that is plus the start and end points of the contour, the putative fin boundary is sampled between every keypoint pair.

$S = \{1, 2, 4, 8\}$, with a constant value of $m = 2$ in Equation 1. This ensures that the scale of the DoG filter used to compute descriptions changes while its shape remains fixed. Each DoG_N descriptor is $L2$ normalised to allow similarities between descriptors to be computed using Euclidean distance. FLANN [28] is employed to store descriptors and to perform efficient approximate nearest neighbour searches. Classification is then performed using each D_Q^j separately and then combined, with each descriptor scale weighted equally ($w_j = 1$).

5 Dataset and Evaluation

In order to benchmark individual fin classification, we used a dataset representing 85 individuals and consisting of 2456 images (see Acknowledgements for data source). For each individual there were on average 29 images (standard deviation of 28). The minimum number for an individual was two. As such, when the dataset was split into labelled and test images, just one labelled training example was selected to represent each shark. The remaining 2371 images were used as queries all of which show at least 25% of the fin’s trailing edge. They exhibited significant variability in waterline and white splash occlusion, view-point, orientation and scale (see Figure 1 and Figure 5 for example images).

Performance Evaluation. Two measures are reported for performance evaluation. Both are based on average precision as the classifier returns a ranked list of candidate identities, each associated with a score as computed according to Equations 2 or 3. The first is average precision (AP) computed for all test images. For the second, we compute precision-recall curves for each individual and then take the mean of the individual average precision scores (mAP). This second measure avoids bias towards individuals with large numbers of test images. In each case, AP is computed as area under precision-recall curves computed directly using the individuals’ scores, in contrast say to the ranking method employed in [16].

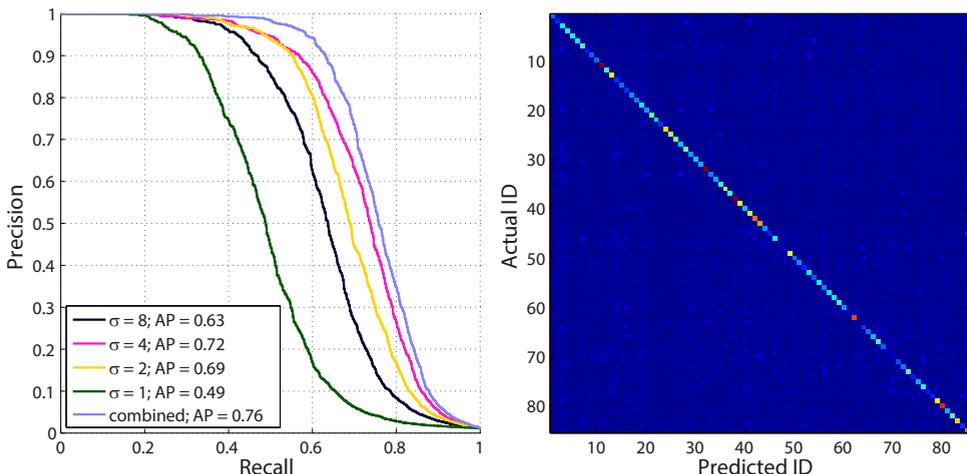


Figure 4: INDIVIDUAL IDENTIFICATION RESULTS. (left) precision-recall curve for individual identification highlights superior performance of multi-scale encoding in blue; (right) confusion matrix of individual matching (ground truth generated by experts from SaveOurSeas Foundation, see Acknowledgements).

6 Individual Identification Results

The mAP and AP scores for individual identification are shown in Table 2. Overall, our contour stroke model for fin detection combined with a combinatorial biometric contour encoding proves suitable for the task of individual fin identification. Of the 2371 query instances presented to the system, 72% are correctly identified as a particular shark with a mAP of 0.79. Figure 5 illustrates examples of fin matches. Not constraining the dataset to good

Sigma	8	4	2	1	combined
mAP	0.67	0.74	0.73	0.56	0.79
AP (all images)	0.63	0.72	0.69	0.49	0.76

Table 2: INDIVIDUAL IDENTIFICATION RESULTS

quality images of distinctive fins comes at a price: that is poor recognition performance for the remaining 28% of the dataset corpus, with 14% of instances being ranked outside the top ten. An examination of recognition performance for high quality fin detections ($F_{\text{inst}} > 0.9$, measured against human labelled ground truth) provides insight into the affect of fin detection on individual identification. Of 217 such detections, where additionally, the entire fin contour was clearly visible, 82% were correctly identified with a mAP of 0.84. In 91% of cases, the correct identity was returned in the top ten ranks. Thus, approximately 9% of fin instances cannot be classified correctly, independent of the quality of the detected contour. The results demonstrate the benefit of combining descriptors computed for independent spectral components of fin shape, as shown by a 6.7% gain in performance compared to that obtained using any individual spectral band alone.

7 Conclusion

We have described a vision framework for automatically identifying individual great white sharks as they appear in unconstrained imagery. To do so we have introduced: 1) a novel contour stroke model that partitions ultrametric contour maps and detects objects based on the resulting open contour descriptions. We have shown that this process simultaneously generates object candidates and separates them from background clutter. 2) A novel spectral and combinatorial method for encoding smooth object boundaries biometrically has been described. The method is both discriminative and robust and shows promising individual shark fin identification performance when employed in a one shot learning paradigm. 3) We have tested performance on a, for the domain, large dataset. We conclude practical applicability at accuracy levels ready to assist human identification efforts without a need for any manual labelling.

As with [8], we expect our framework to generalise to other classes of smooth object, in particular animal species exhibiting biometrically distinctive contours, albeit with new classifiers trained for object class detection.

Acknowledgements

B.H. was supported by EPSRC grant EP/E501214/1. We gratefully acknowledge Michael Scholl and the Save Our Seas Foundation for providing fin images and ground truth labels.

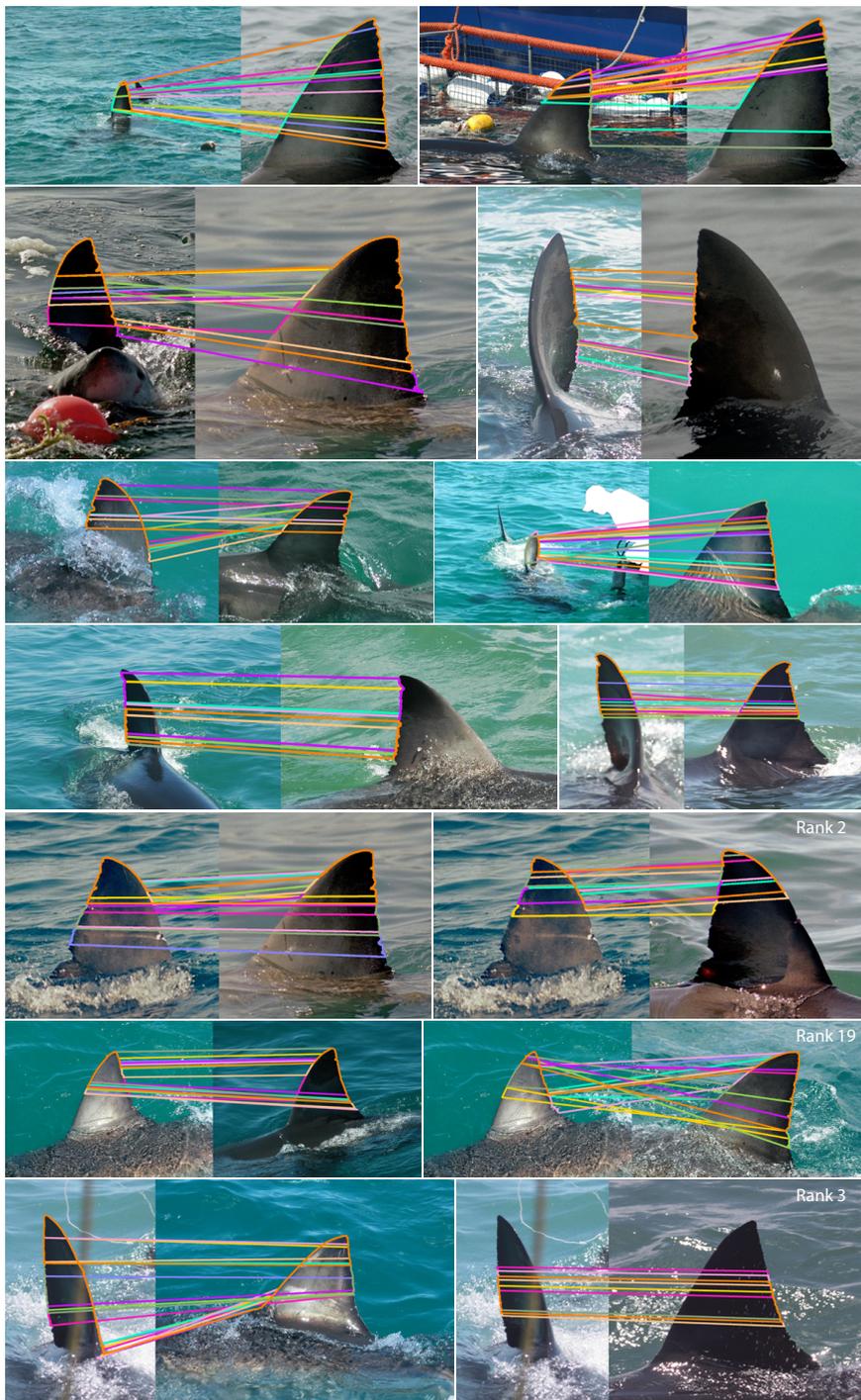


Figure 5: INDIVIDUAL IDENTIFICATION EXAMPLES: left images are queries and right ones are predicted individuals. Coloured lines indicate start and end of the ten sections contributing most evidence for the matched individual. For illustration of false matches, bottom three rows, left pairs, show misidentifications while correct matches are shown right.

References

- [1] Scot D Anderson, Taylor K Chapple, Salvador J Jorgensen, A Peter Klimley, and Barbara A Block. Long-term individual identification and site fidelity of white sharks, carcharodon carcharias, off california using dorsal fins. *Marine Biology*, 158(6):1233–1237, 2011.
- [2] Ognjen Arandjelovic. Object matching using boundary descriptors. In *BMVC 2012: Proceeding of the British Machine Vision Conference 2012*, pages 1–11. BMVA Press, 2012.
- [3] Relja Arandjelovic and Andrew Zisserman. Smooth object retrieval using a bag of boundaries. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 375–382. IEEE, 2011.
- [4] Pablo Arbeláez, Bharath Hariharan, Chunhui Gu, Saurabh Gupta, Lubomir Bourdev, and Jitendra Malik. Semantic segmentation using regions and parts. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3378–3385. IEEE, 2012.
- [5] Pablo Arbeláez, Jordi Pont-Tuset, Jonathan T Barron, Ferran Marques, and Jitendra Malik. Multiscale combinatorial grouping. *CVPR*, 2014.
- [6] Robin W Baird, Antoinette M Gorgone, Daniel J McSweeney, Daniel L Webster, Dan R Salden, Mark H Deakos, Allan D Ligon, Gregory S Schorr, Jay Barlow, and Sabre D Mahaffy. False killer whales (*pseudorca crassidens*) around the main hawaiian islands: Long-term site fidelity, inter-island movements, and association patterns. *Marine Mammal Science*, 24(3):591–612, 2008.
- [7] Robin W Baird, Antoinette M Gorgone, Daniel J McSweeney, Allan D Ligon, Mark H Deakos, Daniel L Webster, Gregory S Schorr, Karen K Martien, Dan R Salden, and Sabre D Mahaffy. Population structure of island-associated dolphins: Evidence from photo-identification of common bottlenose dolphins (*tursiops truncatus*) in the main hawaiian islands. *Marine Mammal Science*, 25(2):251–274, 2009.
- [8] Oren Boiman, Eli Shechtman, and Michal Irani. In defense of nearest-neighbor based image classification. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.
- [9] Ramón Bonfil, Michael Meyer, Michael C Scholl, Ryan Johnson, Shannon O’Brien, Herman Oosthuizen, Stephan Swanson, Deon Kotze, and Michael Paterson. Transoceanic migration, spatial dynamics, and population linkages of white sharks. *Science*, 310(5745):100–103, 2005.
- [10] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [11] Joao Carreira and Cristian Sminchisescu. Constrained parametric min-cuts for automatic object segmentation. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3241–3248. IEEE, 2010.

- [12] Taylor K Chapple, Salvador J Jorgensen, Scot D Anderson, Paul E Kanive, A Peter Klimley, Louis W Botsford, and Barbara A Block. A first estimate of white shark, *carcharodon carcharias*, abundance off central california. *Biology Letters*, 7(4):581–583, 2011.
- [13] Jonathan P Crall, Charles V Stewart, Tanya Y Berger-Wolf, Daniel I Rubenstein, and Siva R Sundaresan. Hotspotter: patterned species instance recognition. In *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*, pages 230–237. IEEE, 2013.
- [14] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.
- [15] Piotr Dollár and C Lawrence Zitnick. Structured forests for fast edge detection. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 1841–1848. IEEE, 2013.
- [16] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136.
- [17] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1627–1645, 2010.
- [18] Jan-Mark Geusebroek, Gertjan J Burghouts, and Arnold WM Smeulders. The amsterdam library of object images. *International Journal of Computer Vision*, 61(1):103–112, 2005.
- [19] Chunhui Gu, Joseph J Lim, Pablo Arbeláez, and Jitendra Malik. Recognition using regions. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1030–1037. IEEE, 2009.
- [20] Gilbert R Hillman, Hemant Tagare, Karin Elder, Alexander Drobyshevski, David Weller, and B Wursig. Shape descriptors computed from photographs of dolphin dorsal fins for use as database indices. In *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, volume 2, pages 970–973. IEEE.
- [21] GR Hillman, GA Gailey, N Kehtarnavaz, A Drobyshevsky, BN Araabi, HD Tagare, and DW Weller. Computer-assisted photo-identification of individual marine vertebrates: a multi-species system. *Aquatic Mammals*, 29(1):117–123, 2003.
- [22] Marcella J Kelly. Computer-aided photograph matching in studies using individual identification: an example from serengeti cheetahs. *Journal of Mammalogy*, 82(2):440–449, 2001.
- [23] Hjalmar S Kühl and Tilo Burghardt. Animal biometrics: quantifying and detecting phenotypic appearance. *Trends in ecology & evolution*, 28(7):432–441, 2013.

- [24] Fuxin Li, Joao Carreira, and Cristian Sminchisescu. Object recognition as ranking holistic figure-ground hypotheses. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1712–1719. IEEE, 2010.
- [25] AD Marshall and SJ Pierce. The use and abuse of photographic identification in sharks and rays. *Journal of fish biology*, 80(5):1361–1379, 2012.
- [26] David R Martin, Charless C Fowlkes, and Jitendra Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(5):530–549, 2004.
- [27] Sancho McCann and David G Lowe. Local naive bayes nearest neighbor for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3650–3656. IEEE, 2012.
- [28] Marius Muja and David G Lowe. Fast approximate nearest neighbors with automatic algorithm configuration.
- [29] Elena Rangelova, Mark Huiskes, and Eric J Pauwels. Towards computer-assisted photo-identification of humpback whales. In *Image Processing, 2004. ICIP'04. 2004 International Conference on*, volume 3, pages 1727–1730. IEEE, 2004.
- [30] Rachel Robbins and Andrew Fox. Further evidence of pigmentation change in white sharks, carcharodon carcharias. *Marine and Freshwater Research*, 63(12):1215–1217, 2013.
- [31] Conrad W Speed, Mark G Meekan, and Corey JA Bradshaw. Spot the match—wildlife photo-identification using information theory. *Frontiers in zoology*, 4(2):1–11, 2007.
- [32] R Stanley. Darwin: Identifying dolphins from dorsal fin images. *Senior Thesis, Eckerd College*, 1995.
- [33] Alison V Towner, Michelle A Wcisel, Ryan R Reisinger, David Edwards, and Oliver JD Jewell. Gauging the threat: the first population estimate for white sharks in south africa using photo identification and automated software. *PloS one*, 8(6):e66035, 2013.
- [34] Jasper RR Uijlings, Koen EA van de Sande, Theo Gevers, and Arnold WM Smeulders. Selective search for object recognition. *International journal of computer vision*, 104(2):154–171, 2013.
- [35] Koen EA Van De Sande, Theo Gevers, and Cees GM Snoek. Evaluating color descriptors for object and scene recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1582–1596, 2010.
- [36] Niccole E Van Hoey. Photo-identification and distribution of bottlenose dolphins (*tursiops truncatus*) off bimini, the bahamas, 2006-2009. 2013.
- [37] AM Van Tienhoven, JE Den Hartog, RA Reijns, and VM Peddemors. A computer-aided program for pattern-matching of natural marks on the spotted raggedtooth shark *carcharias taurus*. *Journal of Applied Ecology*, 44(2):273–280, 2007.

- [38] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511. IEEE, 2001.
- [39] Xiaohong Zhang, Honxing Wang, Mingjian Hong, Ling Xu, Dan Yang, and Brian C Lovell. Robust image corner detection based on scale evolution difference of planar curves. *Pattern Recognition Letters*, 30(4):449–455, 2009.
- [40] Yuanjie Zheng and Chandra Kambhamettu. Learning based digital matting. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 889–896. IEEE, 2009.