Automated Visual Fin Identification of Individual Great White Sharks

Benjamin Hughes and Tilo Burghardt

Abstract This paper discusses the automated visual identification of individual great white sharks from dorsal fin imagery. We propose a computer vision photo ID system and report recognition results over a database of thousands of unconstrained fin images. To the best of our knowledge this line of work establishes the first fully automated contour-based visual ID system in the field of animal biometrics. The approach put forward appreciates shark fins as textureless, flexible and partially occluded objects with an individually characteristic shape. In order to recover animal identities from an image we first introduce an open contour stroke model, which extends multi-scale region segmentation to achieve robust fin detection. Secondly, we show that combinatorial, scale-space selective fingerprinting can successfully encode fin individuality. We then measure the species-specific distribution of visual individuality along the fin contour via an embedding into a global ‘fin space’. Exploiting this domain, we finally propose a non-linear model for individual animal recognition and combine all approaches into a fine-grained multi-instance framework. We provide a system evaluation, compare results to prior work, and report performance and properties in detail.

Keywords animal biometrics, textureless object recognition, shape analysis

1 Introduction

Recognising individuals repeatedly over time is a basic requirement for field-based ecology and related life sciences (Marshall and Pierce, 2012). 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 (Kühl and Burghardt, 2013). To act as an effective aid to biologists, these systems are required to operate reliably on large sets of unconstrained, natural imagery so as to facilitate adoption over widely available, manual or semi-manual identification systems (Stanley, 1995; Van Tienhoven et al, 2007; Ranguelova et al, 2004; Kelly, 2001; Speed et al, 2007). Further automation of identification pipelines for 2D biometric entities is currently subject to extensive research activity (Duyck et al, 2015; Loos and Ernst, 2013; Ravela et al, 2013). Generally, fully automated approaches require at least an integration of 1) a robust fine-grained detection framework to locate the animal or structure of interest in a natural image, and 2) a biometric system to extract individuality-bearing features, normalise and match them (Kühl and Burghardt, 2013). A recent example of such a system for the identification of great apes (Freytag et al, 2016; Loos and Ernst, 2013) uses facial texture information to determine individuals. In fact, all fully automated systems so far rely on the presence of distinctive 2D colour and texture information for object detection as well as biometric analysis.

In this paper we will focus on contour information of textureless objects as biometric entities instead. In specific, we propose a visual identification approach for great white shark fins as schematically outlined in Figure 1, one that extends work in Hughes and Burghardt (2015b) and is applicable to unconstrained fin imagery.

Benjamin Hughes
Dept of Computer Science, University of Bristol
Bristol BS8 1UB, United Kingdom
Now: Save Our Seas Foundation
Rue Philippe Plantamour 20, CH-1201 Geneva, Switzerland
Email: ben@saveourseas.com

Tilo Burghardt
Dept of Computer Science, University of Bristol
Bristol BS8 1UB, United Kingdom
Email: tilo@cs.bris.ac.uk
Fig. 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 embedded into a species-specific ‘fin space’. Shark identities are finally recovered by a non-linear, population-trained identification model that operates on this space.

To the best of our knowledge this line of work establishes the first fully automated contour-based visual ID system in the field of animal biometrics. It automates the pipeline from natural image to animal identity. We build on the fact that fin shape information has been used in the past manually to track individual great white sharks over prolonged periods of time (Anderson et al, 2011) or global space (Bonfil et al, 2005). Shark fin re-identification has also been conducted semi-automatically to support research on the species (Towner et al, 2013; Chapple et al, 2011; Hillman et al, 2003).

We pose the vision task of ‘shark fin identification’ as a fine-grained, multi-instance classification problem for flexible, fairly textureless and possibly partly occluded object parts. ‘Fine-grained’ in that each individual fin, described by a characteristic shape and jagged trailing edge, is a subclass of the parent class great white shark fin. ‘Multi-instance’ since the system must be able to assign multiple semantic labels to an image, each label corresponding to an individual shark present. ‘Flexible’ since fins may bend, and ‘fairly textureless’ since fins lack distinctive 2D texture. In line with work by Arandjelovic and Zisserman (2011), we will also refer to the latter as ‘smooth’. We note that some sharks carry fin pigmentation, yet not all do and its permanence is disputed (Robbins and Fox, 2013). Finally, fin detection poses a part recognition problem since region-based detection of the whole fin would fail to tackle common scenarios: fins are often 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 (top left) and outlines our solution pipeline based on contour information – from image to individual shark ID. We will now review works closest related to the tasks of the recognition pipeline.

2 Related Work and Rationale

Smooth Object Detection. Smooth object detection traditionally builds on utilising boundary and internal contour features, and configurations thereof. Recent approaches (Arandjelovic and Zisserman, 2011; Arandjelovic, 2012) extend these base features by mechanisms for regionalising or globalising information, and infer object presence from learning configuration classifiers. A prominent, recent example is Arandjelovic and Zisserman’s ‘Bag of Boundaries (BoB)’ approach (Arandjelovic and Zisserman, 2011), which employs multi-scale, semi-local shape-based boundary descriptors to regionalise BoB features and predict object presence.

A related, more efficient boundary representation is proposed by Arandjelovic (2012), which focuses on a
Automated Visual Fin Identification of Individual Great White Sharks

Fig. 2 FIN DETECTION AS OPEN CONTOUR STROKES: Multi-scale 2D region-based segmentation algorithms Arbeláez et al (2014) on their own (left images show one level of the ultrametric map) 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).

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 by Arandjelovic (2012) are run on images taken under controlled conditions (Geusebroek et al, 2005), whilst in our work, in common with Arandjelovic and Zisserman (2011), we have the goal of separating objects in natural images and against cluttered backgrounds (see again 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 (Carreira and Sminchisescu, 2010; Li et al, 2010; Uijlings et al, 2013; Gu et al, 2009), or 2) a top-down sliding window detector such as (Viola and Jones, 2001; Dalal and Triggs, 2005; Felzenszwalb et al, 2010) and then segment further detail, or 3) combining the two simultaneously (Arbeláez et al, 2012). We select the first option here since boundary encoding is intrinsic, and bottom-up, efficient and accurate object segmentation has recently become feasible. Arbeláez et al (2014) introduce a fast normalised cuts algorithm, which is used to globalise local edge responses produced by the structured edge detector of Dollár and Zitnick (2013).

However, since fins represent open contour structures (see Figure 2) we require some form of (multi-scale) open contour generation, which is proposed, similar to Arandjelovic (2012), by stipulating key points along closed contours of the ultrametric map as generated by Arbeláez et al (2014). Our proposed contour stroke model (see Section 3) then combines shape information along these open contour sections and nearby regional information to identify and segment fin structures. Note that these are objects which are not present as segments at any level of the ultrametric map.
Biometrics Context. Most closely related within the animal biometrics literature are the computer-assisted fin recognition systems; DARWIN (Stanley, 1995; Stewart et al, 2006) and Finscan (Hillman et al, 2003). DARWIN has been applied to great white sharks (Towner et al, 2013; Chapple et al, 2011) and bottlenose dolphins (Van Hoey, 2013) while Finscan has been applied to false killer whales (Baird et al, 2008), bottlenose dolphins (Baird et al, 2009) and great white sharks, among other species (Hillman et al, 2003). However both differ significantly from our work in that they rely on user interaction to detect and extract fin instances. Their fin descriptors are also 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 Crall et al (2013), this allows images of individuals to be held in tree-based search structures, which facilitate identity discovery in sub-linear time.

Paper Structure. The paper covers six further sections, which will detail the methodology and algorithms proposed, and report on application results and discuss our approach in its wider context. In (3), in accordance with Hughes and Burghardt (2015b), a contour stroke model for fin detection is presented combining a partitioning of ultrametric contour maps with normal encoded shape and nearby opponentSIFT descriptors, whose normal encoded shape and nearby opponentSIFT descriptors feed into a random forest regressor to detect fin objects.

Stage 1: Hierarchical Segmentation. We use work by Arbeláez et al (2014) to generate a region hierarchy in the form of an ultrametric map. This provides sets of closed contours 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 Carreira and Sminchisescu (2010), we then employ region rejection to remove areas too small to represent a fin, or too similar to another region. We subsequently
rank remaining regions, again by their location in the hierarchy, and retain the top \( k \) regions, choosing \( k = 12 \) empirically for the results given in this paper.

**Stage 2: Generating Fin Candidates.** In almost all cases, the segmentation produces at least one single region, within the 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 (see Figure 2). 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 thereby stroke generation. We detect boundary keypoints using the Difference of Gaussian (DoG) corner detector of Zhang et al (2009). 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 \( D(u, \sigma) \):

\[
D(u, \sigma) = \left[ DoG * x(u) \right]^2 + \left[ DoG * y(u) \right]^2
\]

\[
= [G(u, m\sigma) * x(u) - G(u, \sigma) * x(u)]^2 +
[G(u, m\sigma) * y(u) - G(u, \sigma) * y(u)]^2
\]

where \( G(u, \sigma) \) is a zero mean Gaussian function with standard deviation \( \sigma \), and \( m > 0 \) is a multiplication factor. Viewed as a bandpass filter, by varying \( m \) and \( \sigma \), 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 (see Figure 4). This allows for the selection of the \( n \) peaks with largest prominence suppressing other, locally non-maximal corner responses. Choosing small values of \( \sigma \) ensures accurate keypoint localisation whilst 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 without taking the two encoding directions (clockwise and anticlockwise) into account. 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 \( F^g_{\text{inst}} \). Evaluated with respect to a human-labelled ground truth contour, we use the standard \( F^g \)-measure for evaluating contour detections based on bipartite matching of boundary pixels (Martin et al, 2004). 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 \( F^g_{\text{inst}} = 0.97 \) measured against human-labelled ground truth contours for 240 randomly selected images. By means of comparison, the average quality of the pool of \( k = 12 \) closed region contours is \( F^g_{\text{inst}} = 0.75 \).

**Stage 3: Fin Candidate Scoring.** For training and testing the candidate classifier, 240 high visibility (H) 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. In addition, we perform validation using a second set of 165 ‘lower’ visibility (L) images where fins are partially occluded, again, selected at random. This will enable us to establish whether the trained model is representative given partial occlusion. Examples of each image type are shown in Figure 5.

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 180-dimensional feature vector consisting of two components, contributing 2D and 1D distinctive information, respectively.
The first is a bag of opponentSIFT (Van De Sande et al, 2010) 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. This descriptor is utilised to capture 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. Note that the opponentSIFT histogram is independent of encoding direction whilst the histogram of boundary normals is dependent on it\(^2\).

In either case, the two components are \(L_2\) normalised and concatenated to produce the final descriptor. A random forest regressor (Breiman, 2001) is 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 (Martin et al, 2004). 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 resulting detector then produces a set of candidate detections, each with a predicted quality score \(F_{p_{\text{inst}}}\). Figure 6 illustrates example candidates together with their scores.

Measuring Detection Performance. We use 1) average precision (AP\(_t\)\(_{\text{det}}\)), the area under the precision-recall (PR) curve for a given threshold \(t\), and 2) volume under PR surface (AP\(_{\text{vol}}\)) as evaluation metrics. In order to generalise AP\(_t\)\(_{\text{det}}\), the AP\(_{\text{vol}}\) measure was proposed by Hariharan et al (2014) for simultaneous object detection and segmentation. It measures the vol-

\(^2\) When training the histogram of boundary model, we flip images so the fin is facing the same way in each. For testing, we compute two feature vectors, one for each fin direction. We then obtain a predicted quality score for each direction and take the maximum over directions as the predicted quality for that stroke.
**Fig. 7** FIN DETECTION RESULTS: (A,B) Scatter plots show that the full fin detection model is able strongly to predict, as captured by $F^p_{\text{inst}}$, the true quality of fin candidates $F^g_{\text{inst}}$ for both high and low visibility images. (C) The plot summarises performance at different stages of fin detection. Note, that for the ‘segmentation’ line, $AP^t_{\text{det}}$ is equivalent to the proportion of fins for which it is possible to obtain a stroke of quality $F^g_{\text{inst}} \geq t$, given a machine generated segmentation. (D) The plot shows PR plots for both high and low visibility images at different thresholds.

We base our fin detection evaluation on AP instead of a receiver operating characteristic (ROC) curve based measure such as AUC-ROC, since the choice of precision over FPR increases evaluation sensitivity to changing numbers of false positives in the presence of large numbers of negative examples (Davis and Goadrich, 2006). In addition, the choice of AP-based evaluation is in line, not only with Hariharan et al (2014), but also with the methodology adopted in the object detection components of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al, 2015) and in the PASCAL Visual Object Challenge (PASCAL VOC) (Everingham et al, 2010), two standard benchmarks for visual object recognition.

**Fin Detection Results.** Results for fin candidate generation and detection are shown in Figure 7, Table 1 and Table 2. Scatter plots in Figure 7 for high and lower visibility images confirm that the model is able strongly to identify fins, and many high quality candidates are generated as shown by the large number of instances with high $F^g_{\text{inst}}$ scores. The Pearson correlation coefficients between true and predicted quality scores are 0.95 and 0.93, respectively.

The plot of Figure 7(C) summarises performance at different stages of fin detection. We note that for segmentation, a stroke of quality $F^g_{\text{inst}} \geq 0.95$ is possible for almost all fin instances (98.3%), with an average achievable quality, $AP^{\text{vol}}$, of 0.99. Candidate generation also performs well. It can be seen that for almost all high visibility fins (98.3%), a candidate of $F^g_{\text{inst}} > 0.9$ is generated and $F^g_{\text{inst}} > 0.85$ for 98.8% of lower visibility fins. Across all thresholds and fin instances, average achievable qualities of 0.97 and 0.96 are seen respectively. Table 1 summarises these intermediate results.

Finally, we show results for the whole pipeline in Figure 7(D) and Table 2, that of combined segmentation, candidate generation and candidate classification. Here we see that a candidate of quality $F^g_{\text{inst}} \geq 0.83$ is generated and recognised (with $AP^t_{\text{det}} = 0.98$) for almost all high visibility fins ($F^g_{\text{inst}} \geq 0.78$ for lower visibility with $AP^t_{\text{det}} = 0.99$), as indicated by $AP^t_{\text{det}}$ close to 1 for these quality thresholds, with $AP^t_{\text{det}} = 1$ only possible if both $p^t_{\text{det}} = 1$ and $R^t_{\text{det}} = 1$.

To fully understand values of $AP^t_{\text{det}} < 1$, we must consider detection precision and recall separately, as shown in Figure 7(D). Here we show PR curves for

| Table 1 INTERMEDIATE RESULTS ($AP^t_{\text{det}}$) |
|-----------------------------------------------|
| $t=0.7$ | $t=0.85$ | $t=0.9$ |
| Segmentation | 1.0 | 0.99 | 0.99 | 0.99 |
| Candidate gen. (H) | 0.99 | 0.98 | 0.98 | 0.97 |
| Candidate gen. (L) | 1.0 | 0.99 | 0.92 | 0.96 |

| Table 2 FIN DETECTION RESULTS ($AP^t_{\text{det}}$) |
|-----------------------------------------------|
| Feature type | $t=0.7$ | $t=0.85$ | $t=0.9$ | $AP^{\text{vol}}$ |
| High Visibility (H) |
| OpponentSIFT | 0.99 | 0.85 | 0.73 | - |
| Normal | 0.98 | 0.85 | 0.7 | - |
| Combined | 0.98 | 0.95 | 0.86 | 0.86 |
| Lower Visibility (L) |
| Combined | 1.0 | 0.93 | 0.62 | 0.89 |
selected quality thresholds of the complete detection pipeline. We see for example that for \( t = 0.85 \), perfect precision (\( P_{\text{det}} = 1.0 \)) is achieved for about 63\% of both high and lower visibility fins (\( R_{\text{det}}^t = 0.63 \)), after which, false positives are introduced as shown by reduced precision. We also see that detection recall does not equal 1 for any value of precision, repeating the observation that a candidate of this quality is not generated for every fin. Meanwhile, we see near perfect detection if we accept candidates with \( P^0_{\text{inst}} \geq 0.7 \).

Finally, observing the summary of results in Table 2, we see the effectiveness of the different features types for fin candidate classification. It can be seen that while both opponentSIFT and normal features enable good detection performance (say for \( t = 0.7 \)), a combination of the two is required to obtain good recognition of the highest quality candidates at \( t = 0.9 \). In summary, for almost all fin instances, a high quality candidate is generated and recognised with high precision, demonstrating the effectiveness of our contour stroke model for the task at hand.

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 whilst being robust to noisy, partially occluded input generated by automatic shape extraction.

Global shape descriptions, as used in Stewman et al (2006), maximise inter-class variance but are sensitive to partial occlusions and object-contour detection errors, while the removal of misalignment variables such as in- and out-of-plane rotation rely upon computing point correspondences and inefficient pairwise matching.

By contrast, the semi-local descriptions of Arandjelovic and Zisserman (2011); Arandjelovic (2012) are robust and allow efficient matching, but their encoding of inter-class variance will always be sub-optimal. To maximise the descriptiveness of features, we utilise both semi-local and global shape descriptions with a framework extending that used to generate fin candidates.

Edge Refinement. Our segmentation and contour partitioning framework so far 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 (Anderson et al, 2011; Bonfil et al, 2005). 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 and Kumbhamettu (2009). This produces an opacity mask \( \alpha \) which defines a soft segmentation of the image (\( a_i \in [0,1] \)). 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 as a precursor to biometric encoding. Full details of this edge refinement procedure can be found in Hughes and Burghardt (2015a).

**Generating Boundary Subsections.** As a first step towards a biometric encoding, we detect salient boundary keypoints on the extracted contour strokes to produce stably recognisable contour subsections that serve as descriptor regions. For keypoint detection we use the same approach as that used for detecting keypoints when generating fin candidates, as described in Section 3. To generate boundary subsections, we resample fin candidates to a fixed resolution of 1024 pixels and compute \( D(u, \sigma) \) in Equation 1, re-parameterised with \( \sigma = 2 \) and \( m = 8 \). Subdivision by these keypoints yields \( \binom{n}{2} = 1225 \) contour subsections\(^3\). Note that for reference images, we encode subsections in both directions. For test images, we encode in one direction only. As a result, later subsection matching does not need to consider the directions. The approach is illustrated in Figure 8.

**Boundary Descriptors.** Following the generation of boundary subsections, the task is to encode their shape information. We investigate two regimes for subsection description: the standard DoG norm (DoG\(_N\)) as defined in Equation 1, and the boundary descriptor of Arandjelovic (2012). DoG\(_N\) provides a number of properties relevant to biometric contour encoding: first, the associated metric is suitable for establishing similarity between descriptors, meaning contour sections can be matched efficiently. Secondly, by varying the parameters \( \sigma \) and \( m \), the description can be tuned to encode different components of shape scale-space. Third, the descriptor is rotation invariant and robust to changes in viewpoint (see Figure 16).

We also consider the boundary descriptor of Arandjelovic (2012) composed of a vector of boundary normals, denoted \( \mathcal{N}(u, \sigma) \). At each vertex the normal vector of the contour is computed and the two orthogonal components are concatenated to yield the descriptor:

\[
\mathcal{N}(u, \sigma) = (G(u, \sigma) \ast x(u), G(u, \sigma) \ast y(u)) \tag{2}
\]

This normal descriptor lacks rotational invariance. This is overcome by aligning the ends of each subsection with a fixed axis as a precursor to descriptor computation.

\(^3\) Taking 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.
As illustrated in Figure 9 over the entire fin segment, both DoG\textsubscript{N} and Arandjelovic’s normal descriptor provide spatial and scale selectivity.

5 Identification Baseline via LNBNN

As noted by Boiman et al (2008), 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 (McCann and Lowe, 2012), and similar to that employed by Crall et al (2013) in the context of patterned species individual identification, to provide a recognition baseline.

**Implementation Details.** To achieve scale normalisation, each contour subsection is re-sampled to a fixed length of 256 pixels. DoG\textsubscript{N} and normal descriptors are computed at filter scales \( S = \{ 1, 2, 4, 8 \} \), with a constant value of \( m = 2 \) in the DoG\textsubscript{N} case. Each descriptor is \( L2 \) normalised to allow similarities between descriptors to be computed using Euclidean distance. FLANN (Muja and Lowe, 2009) is employed to store descriptors and to perform efficient approximate nearest neighbour searches. Classification is performed at each scale separately for both descriptor types and then combined, with each scale weighted equally (\( w_j = 1 \)).

**Dataset.** In order to benchmark individual fin classification, we use a dataset representing 85 individuals and consisting of 2456 images (see Acknowledgements...
Table 3 INDIVIDUAL LNBNN ID RESULTS

| Encoding  | σ = 8 | σ = 4 | σ = 2 | σ = 1 | combined |
|-----------|-------|-------|-------|-------|---------|
| AP:DoG_N  | 0.63  | 0.72  | 0.69  | 0.49  | 0.76    |
| AP:Norm   | 0.33  | 0.70  | 0.72  | 0.65  | 0.72    |
| mAP:DoG_N | 0.67  | 0.74  | 0.73  | 0.56  | 0.79    |
| mAP:Norm  | 0.49  | 0.75  | 0.76  | 0.73  | 0.76    |

2 training images per class: 2286 queries

| Encoding  | AP  | mAP  |
|-----------|-----|------|
| AP:SIFT   | 0.20| -    |
| mAP:SIFT  | 0.35| -    |
| AP:DoG_N | 0.81| 0.83 |
| mAP:DoG_N| 0.83| -    |

for data source). For each individual there are 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, viewpoint, orientation and scale (see Figure 1 and Figure 16 for example images).

Performance Evaluation Measures. 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 3 or 5. The first is AP, computed for all test images. For the second, we compute AP for each individual and then take the mean of the individual AP 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 employed in Everingham et al. (2014).

Results. The mAP and AP scores for DoG_N and normal-based individual identification are shown in Table 3. Overall, our contour stroke model for fin detection combined with a combinatorial biometric contour encoding proves suitable for the task of individual fin identification. For DoG_N, as reported in Hughes and Burghardt (2015b) for one-shot-learning, of the 2371 query instances presented to the system, a particular shark is 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 could not be classified correctly, independent of the quality of the detected contour.

The results demonstrate the benefit of combining DoG_N descriptors computed for independent scale-space components of fin shape, as shown by a 6.7% gain in AP performance from AP=0.72 to AP=0.76 compared to that obtained using any individual scale alone.

The normal encoding also proves suitable for individual recognition, with AP of 0.72 and mAP of 0.76, although the best performance obtained with this descriptor type falls below the multi-scale DoG_N approach.

Finally, we observe that the DoG_N and normal approaches produce different predictions on a significant set of samples, pointing towards an opportunity in combining these classifiers, depending on fin structure. This complementarity is exploited in Section 6.

Fig. 10 PRECISION-RECALL CURVES FOR LNBNN. Precision-recall curves for DoG_N and normal fin encodings, comparing identification via whole fins and just trailing edges.
Comparison with Off-the-shelf Features. To put the performance of our biometric contour representation in context, we report individual fin identification results using a methodology previously applied to patterned species individual recognition (Crall et al., 2013). In our case, a sparse, affine covariant SIFT encoding (Mikolajczyk and Schmid, 2004) of fin shape and surface texture is generated by detecting features centred within closed regions, created by drawing straight lines between the two ends of detected fin stokes (illustrated using dashed lines in Figure 2). As before, LNBNN (Equations 3 and 4) is used for individual classification. In this experiment (and only this experiment) two training images are used per individual, one for each side of the fin, leaving 2286 query images.

Results in Table 3 unambiguously demonstrate the superiority of our biometric contour representation over one describing surface texture, for individual fin identification. Using SIFT features, fins are identified with mAP of 0.35 (AP=0.2). Using exactly the same training data, this compares with mAP of 0.83 using the combinatorial multi-scale DoG encoding (AP=0.81). Interestingly however, 45 fin instances, misclassified using biometric contour encoding, are correctly identified using SIFT, with examples shown in Figure 11. Noting that the permanence of fin surface markings additionally captured by 2D features such as SIFT is disputed (Robbins and Fox, 2013), this observation nevertheless suggests that texture-based representations may have potential utility, at least for a sub-set of the population and over short observation windows.

6 Construction of a Population-wide Fin Space

In this section, we introduce a globally normalised cross-class (cross-individual) coordinate system over both descriptors DoG\textsubscript{N} and normals, i.e. a global ‘fin space’, in which we embed fin descriptors along the dimensions of descriptor type, spatial location and spatial extent on the fin contour, as well as along feature scale. The resulting 4D fin space is illustrated in Figure 12.

This space allows for reasoning about and learning of population-wide properties using anatomically interpretable dimensions; be that to 1) quantify the distinctiveness of feature descriptors by their type, location or extent on the fin, or to 2) move from a non-parametric and linear method of cue combination to one that non-linearly learns how to combine indexed evidence from across the fin space. Importantly, this entails learning a single model for the species, one which can be seen as characterising a species-specific pattern of individual distinctiveness, and not one that learns a pattern of uniqueness solely for any given individual.

![Fig. 11 EXAMPLE IDENTIFICATIONS USING AFFINE-COVARIANT SIFT DESCRIPTIONS](image)

**Fig. 11** EXAMPLE IDENTIFICATIONS USING AFFINE-COVARIANT SIFT DESCRIPTIONS: Rarely, fins misclassified using biometric contour representations are correctly identified using surface texture descriptors. Here, two such examples are shown, with query images on the left of each pair. The coloured lines represent discriminative feature matches (as evaluated by the value of \(f(d,c)\) in Equation 4)

Embedding Descriptors into Fin Space. The fabric of the proposed fin space can be described as subdividing the leading and trailing edges of fins into \((n=5)\) equally sized partitions\(^4\). We then consider every connected combination of partitions yielding 55 spatial bins for each of the two feature types.

As illustrated in Figure 13, fin subsections can be mapped to spatial bins by first assigning them to partitions - a subsection is said to occupy a partition if it occupies more than half of it. Finally, each subsection is assigned to the spatial bin that corresponds to the set of partitions it occupies. Scale-space partitioning is achieved by dividing filter scale into five bins.

More formally, this yields an overall set of bins given by \(B = \{(\sigma_1^a, \sigma_2^a), \ldots, (\sigma_k^a, \sigma_{k+1}^a), \ldots, (\sigma_{m-1}^a, \sigma_m^a)\}\) and the set of filter scales is \(S^a = \{\sigma_1^a, \ldots, \sigma_k^a, \ldots, \sigma_m^a\}\). Here \(g\) denotes that filter scale is considered as a proportion of the fin contour length globally.

Defined globally, the filter scale of the \(i^{th}\) subsection descriptor computed at scale \(j\) (as in Equation 5) can be expressed as \(\sigma_j^i = \sigma_j / \ell_n \cdot p\) where \(p\) expresses the length of the subsection as a proportion of the length of the fin contour, and \(\ell_n\) is the number of samples used to encode

\(^4\) As the lengths of either edge of the fin are not necessarily the same, the size of the partitions on the leading edge are not necessarily the same as those on the trailing edge.
the subsection. Having computed $\sigma_{i,j}^g$, the descriptor is mapped to the corresponding bin.

7 Non-Linear Model Exploiting Fin Space

In this section we show that learning distributions of reliable match locations in fin space can significantly improve identification rates compared to the baseline. This appreciates the fact that certain feature combinations in fin space are common and not individually discriminative in sharks, whilst others are highly distinctive. To implement a practical approach that captures such species-specific information, we learn a non-linear map from patterns of matching locations in fin space to likelihoods of reliability for identification.

Obtaining Scoring Vectors from Fin Space. As in the baseline case, for each query descriptor (representing the input image) and for each class (representing the individuals), we find the nearest reference descriptor in that class, i.e. perform max-pooling over the class. As described in Section 5, based on the distance to that nearest neighbour and the distance to the nearest neighbour in another class, we compute a local match score according to Equation 4.

Now, instead of sum-pooling local match scores over class labels directly, as performed in Equations 3 and 5, we first project local match scores into fin space via their associated reference descriptors, and then perform sum-pooling over fin space bins (see Figure 14). As a result, for each class and for each discrete fin space location, we obtain a score. These scores form a vector of dimensionality equal to the cardinality of fin space. As such, each query-class comparison yields such a vector.

Learning a Non-Linear Identification Model. The described procedure rearranges matching information so that the scoring pattern as observed spatially and in scale-space along the fin, as well as over descriptor types, is made explicit by the scoring vector. We now analyse the structure of scoring vectors over an entire population of fins to learn and predict their reliability for inferring animal identity. This procedure is designed to selectively combine descriptor evidence (see Section 5), exploit the observed variance in local distinctiveness (see Figure 12), and address potential correlations between features in fin space. To allow for complex, non-linear relationships between scoring structures and identification reliability, we select a random forest classifier to implement the mapping.

Practically, we train the random forest to map from query-class scoring vectors to probability distributions over binary match category labels ‘query-is-same-class’ and ‘query-is-not-same-class’. Importantly, performing two-fold cross-validation, the dataset is split randomly by individual, and not by query, when training and evaluating the classifier. This ensures that what is learned generalises across the species and does not over-fit the individuals in the present dataset.

Final Results. Evaluation is performed by reporting AP and precision-recall curves over the same 2371 queries as used to obtain the identification baselines in Section 5. We present the results in Figure 15. It can be seen that, overall, the final fin space approach achieves an AP of 0.81, representing 7% and 12% performance gains over the DoG_N and normal baselines, respectively. The results also clearly demonstrate the benefit of selectively combining both descriptor types - precision measures are improved or kept across the entire recall spectrum for a combined, dual descriptor approach.
Local similarities

Max pooling over class

Sum pooling over class

Classification

Query descriptors

Reference descriptors

Max pooling

over class

Sum pooling over class

Score

Query-class match

probability

Identity

Fig. 14 COMPARISON OF BASELINE (top) AND FIN-SPACE IDENTIFICATION SCHEME (bottom). The two paradigms are illustrated proceeding from left to right. By associating descriptor matching scores (left column) to reference locations in a global fin space (colouration), the improved scheme (bottom) accumulates information not into a single, class-specific scalar (top approach), but forms a scoring vector that encodes the pattern of matchings over fin space. Identity is then judged via a random forest based on the learned reliability of the matching patterns.

8 Conclusions and Future Work

We have described a vision framework for automatically identifying individual great white sharks as they appear in unconstrained imagery as used by white shark researchers. To do so, we have first described a contour stroke model that partitions ultrametric contour maps and detects fin objects based on the resulting open contour descriptions. We have shown that this process simultaneously generates fin object candidates and separates them from background clutter.

Secondly, a multi-scale and combinatorial method for encoding smooth object boundaries biometrically has been described. In combination with an LNBNN classifier, the method is both discriminative and robust, and shows individual shark fin identification performance at a level of AP=0.76 when employed using a multi-scale DoG descriptor in a one shot learning paradigm.

Thirdly, we have introduced a domain-specific ‘fin space’ which indexes fin shapes spatially, by filter scale and along descriptor types. We have measured the distinctiveness for individual shark identification of different regions in this space, providing some insight into the distribution of individuality over the fin.

Finally, we have proposed a shark fin identification framework that achieves an AP=0.81 outperforming the baseline system published in Hughes and Burghardt (2015b). In essence, we achieved this improvement by introducing a non-linear recognition model, which integrates different descriptors and operates based on a population-wide, learned model for predicting identification reliability from matching patterns in fin space.

For the species at hand, we conclude practical applicability at accuracy levels ready to assist human identification efforts without a need for any manual labelling. The approach may therefore be integrated to enhance large scale citizen science (Simpson et al, 2013; Berger-Wolf et al, 2015; Duyck et al, 2015) for ecological data collection of white sharks. A related project to make available this work to the biological research community is underway (Scholl, 2016).

Furthermore, we expect our framework to generalise to other classes of smooth biometric entity, in particular marine life exhibiting individually distinctive fin and fluke contours such as various other species of shark and whale, e.g. humpback whales (Ranguelova et al, 2004).

Dataset

The dataset "FinsScholl2456" containing 2456 images of great white sharks and their IDs was used in this paper. Since the authors and host institution hold no copyright, to obtain a copy please directly contact: Michael C. Scholl, Save Our Seas Foundation (CEO), Rue Philippe Plantamour 20, CH-1201, Geneva, Switzerland; Email: Michael@SaveOurSeas.com
Fig. 16 LNBNN 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. All example matches are obtained using multiscale DoG\textsubscript{N} descriptors combined using the LNBNN classifier. Out of respect for their privacy, the human subject appearing in row 3, column 2, was masked out of the image prior to publication, but only after fin detection and photo-identification results had been obtained.
Acknowledgements

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

References

Anderson SD, Chapple TK, Jorgensen SJ, Klimley AP, Block BA (2011) Long-term individual identification and site fidelity of white sharks, carcharodon carcharias, off California using dorsal fins. Marine Biology 158(6):1233–1237

Arandjelovic O (2012) Object matching using boundary descriptors. In: BMVC 2012: Proceeding of the British Machine Vision Conference 2012, BMVA Press, pp 1–11

Arandjelovic R, Zisserman A (2011) Smooth object retrieval using a bag of boundaries. In: Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE, pp 375–382

Arbeláez P, Hariharan B, Gu C, Gupta S, Bourdev L, Malik J (2012) Semantic segmentation using regions and parts. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, pp 3378–3385

Arbeláez P, Pont-Tuset J, Barron JT, Marques F, Malik J (2014) Multiscale combinatorial grouping. CVPR

Baird RW, Gorgone AM, McSweeney DJ, Webster DL, Salden DR, Deakos MH, Ligon AD, Schorr GS, Barlow J, Mahaffy SD (2008) 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

Baird RW, Gorgone AM, McSweeney DJ, Ligon AD, Deakos MH, Webster DL, Schorr GS, Martien KK, Salden DR, Mahaffy SD (2008) 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

Berger-Wolf TY, Rubenstein DI, Stewart CV, Holmberg J, Parham J, Crall J (2015) Ibeis: Image-based ecological information system: from pixels to science and conservation. The Data for Good Exchange

Boiman O, Shechtman E, Irani M (2008) In defense of nearest-neighbor based image classification. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, IEEE, pp 1–8

Bonfil R, Meýer M, Scholl MC, Johnson R, O’Brien S, Oosthuizen H, Swanson S, Kotze D, Paterson M (2005) Transoceanic migration, spatial dynam-
Gu C, Lim JJ, Arbeláez P, Malik J (2009) Recognition using regions. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, pp 1030–1037

Hariharan B, Arbeláez P, Girshick R, Malik J (2014) Simultaneous detection and segmentation. In: Computer Vision–ECCV 2014, Springer, pp 297–312

Hughes B, Burghardt T (2015a) Affinity matting for pixel-accurate fin shape recovery from great white shark imagery. Machine Vision of Animals and their Behaviour (MVAB), Workshop at BMVC pp 8.1–8.8

Hughes B, Burghardt T (2015b) Automated identification of individual great white sharks from unrestricted fin imagery. 24th British Machine Vision Conference, (BMVC) pp 92.1–92.14

Kelly MJ (2001) Computer-aided photograph matching in studies using individual identification: an example from serengeti cheetahs. Journal of Mammalogy 82(2):440–449

Kühl HS, Burghardt T (2013) Animal biometrics: quantifying and detecting phenotypic appearance. Trends in ecology & evolution 28(7):432–441

Li F, Carreira J, Sminchisescu C (2010) Object recognition as ranking holistic figure-ground hypotheses. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE, pp 1712–1719

Loos A, Ernst A (2013) An automated chimpanzee identification system using face detection and recognition. EURASIP Journal on Image and Video Processing DOI: 10.1186/1687-5281-2013-49

Marshall A, Pierce S (2012) The use and abuse of photographic identification in sharks and rays. Journal of fish biology 80(5):1361–1379

Martin DR, Fowlkes CC, Malik J (2004) 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

McCann S, Lowe DG (2012) Local naive bayes nearest neighbor for image classification. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, pp 3650–3656

Mikolajczyk K, Schmid C (2004) Scale & affine invariant interest point detectors. International journal of computer vision 60(1):63–86

Muja M, Lowe DG (2009) Fast approximate nearest neighbors with automatic algorithm configuration. VISAPP 1(2)

Rangelova E, Huiskes M, Puwela EJ (2004) Towards computer-assisted photo-identification of humpback whales. In: Image Processing, 2004. ICIP’04. 2004 International Conference on, IEEE, vol 3, pp 1727–1730

Ravela S, Duyck J, Finn C (2013) Vision-based biome- metrics for conservation. In: Carrasco-Ochoa, JA, Martínez-Trinidad, JF, Rodríguez, JS, di Baja, GS (eds) MCPR 2012 LNCS 7914:10–19

Robbins R, Fox A (2013) Further evidence of pigmentation change in white sharks, carcharodon carcharias. Marine and Freshwater Research 63(12)

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, et al (2015) Imagenet large scale visual recognition challenge. International Journal of Computer Vision 115(3):211–252

Scholl M (2016) High tech fin recognition. Save Our Seas Magazine http://www.saveourseasmagazine.com/high-tech-fin-recognition

Simpson R, Page KR, Roure DD (2013) Zooniverse: observing the world’s largest citizen science platform. Proceedings of the 23rd International Conference on World Wide Web pp 1049–1054

Speed CW, Meelan MG, Bradshaw CJ (2007) Spot the match–wildlife photo-identification using information theory. Frontiers in zoology 4(2):1–11

Stanley R (1995) Darwin: Identifying dolphins from dorsal fin images. Senior Thesis, Eckerd College

Stewman J, Deubre K, Hale S, Russell A (2006) Iterative 3-d pose correction and content-based image retrieval for dorsal fin recognition. In: Image Analysis and Recognition, Springer, pp 648–660

Towner AV, Wcislo MA, Reisinger RR, Edwards D, Jewell OJ (2013) Gauging the threat: the first population estimate for white sharks in south africa using photo identification and automated software. PloS one 8(6):e66,035

Uijlings JR, van de Sande KE, Gevers T, Smeulders AW (2013) Selective search for object recognition. International journal of computer vision 104(2):154–171

Van De Sande KE, Gevers T, Snoek CG (2010) Evaluating color descriptors for object and scene recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on 32(9):1582–1596

Van Hoey NE (2013) Photo-identification and distribution of bottlenose dolphins (tursiops truncatus) off bimini, the bahamas, 2006-2009
Van Tienhoven A, Den Hartog J, Reijns R, Peddemors V (2007) A computer-aided program for pattern-matching of natural marks on the spotted ragged-tooth shark carcharias taurus. Journal of Applied Ecology 44(2):273–280
Viola P, Jones M (2001) 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, IEEE, vol 1, pp I–511
Zhang X, Wang H, Hong M, Xu L, Yang D, Lovell BC (2009) Robust image corner detection based on scale evolution difference of planar curves. Pattern Recognition Letters 30(4):449–455
Zheng Y, Kambhamettu C (2009) Learning based digital matting. In: Computer Vision, 2009 IEEE 12th International Conference on, IEEE, pp 889–896