CrossMark

# Automated Visual Fin Identification of Individual Great White Sharks

**Benjamin Hughes**[1,2] · **Tilo Burghardt**[2]

**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.

✉ Benjamin Hughes
ben@saveourseas.com

Tilo Burghardt
tilo@cs.bris.ac.uk

[1] Save Our Seas Foundation, Rue Philippe Plantamour 20, CH-1201 Geneva, Switzerland

[2] Department of Computer Science, University of Bristol, Bristol BS8 1UB, UK

## 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; 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
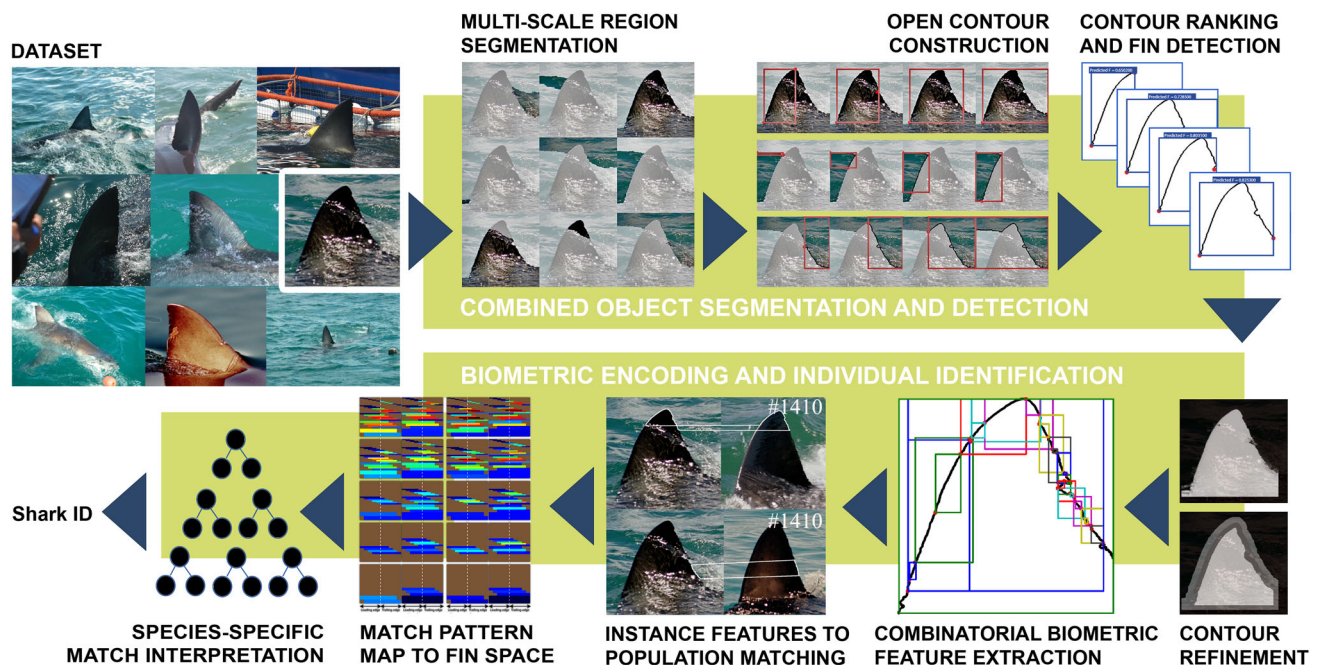
**Fig. 1** System overview: We perform a coarse and a fine-grained recognition task. The first is to simultaneously segment and detect shark fins, and the second is to recognise individuals. We begin by segmenting an image into an ultrametric contour map, before partitioning boundaries into sets of open contours. We then train a random forest to rank contours and detect fin candidates based on normal informa- tion 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

fins as schematically outlined in Fig. 1, one that extends work in Hughes and Burghardt (2015b) and is applicable to unconstrained fin imagery. 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 dis-

puted (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

### 2.1 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 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 by Arandjelovic (2012) are run on images taken under controlled conditions (Geusebroek et al. 2005), whilst in our work, in common with arandjelovic11, we have the goal of separating objects in natural images and against cluttered backgrounds (see again Figure 1).

### 2.2 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 Fig. 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 Sect. 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.

### 2.3 Biometrics Context

Most closely related within the animal biometrics literature are the computer-assisted fin recognition systems; DARWIN (Stanley 1995; Stewman 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 signifi-

cantly 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 Sect. 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.

### 2.4 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 descriptors and dense local features. Then, expanding on previous work, in (4) and (5) a dual biometric encoding scheme for fins and an associated LNBNN baseline identification approach are discussed. In (6) we quantify species-specific visual individuality via a 'fin space', and in (7) an improved non-linear identification framework that uses this space is shown and evaluated. Finally, in (8) we discuss the scope and conclusions of individually identifying great white sharks visually.

## 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) we 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 (Sande et al. 2010). Figure 3 illustrates this fin detection approach in detail.

### 3.1 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
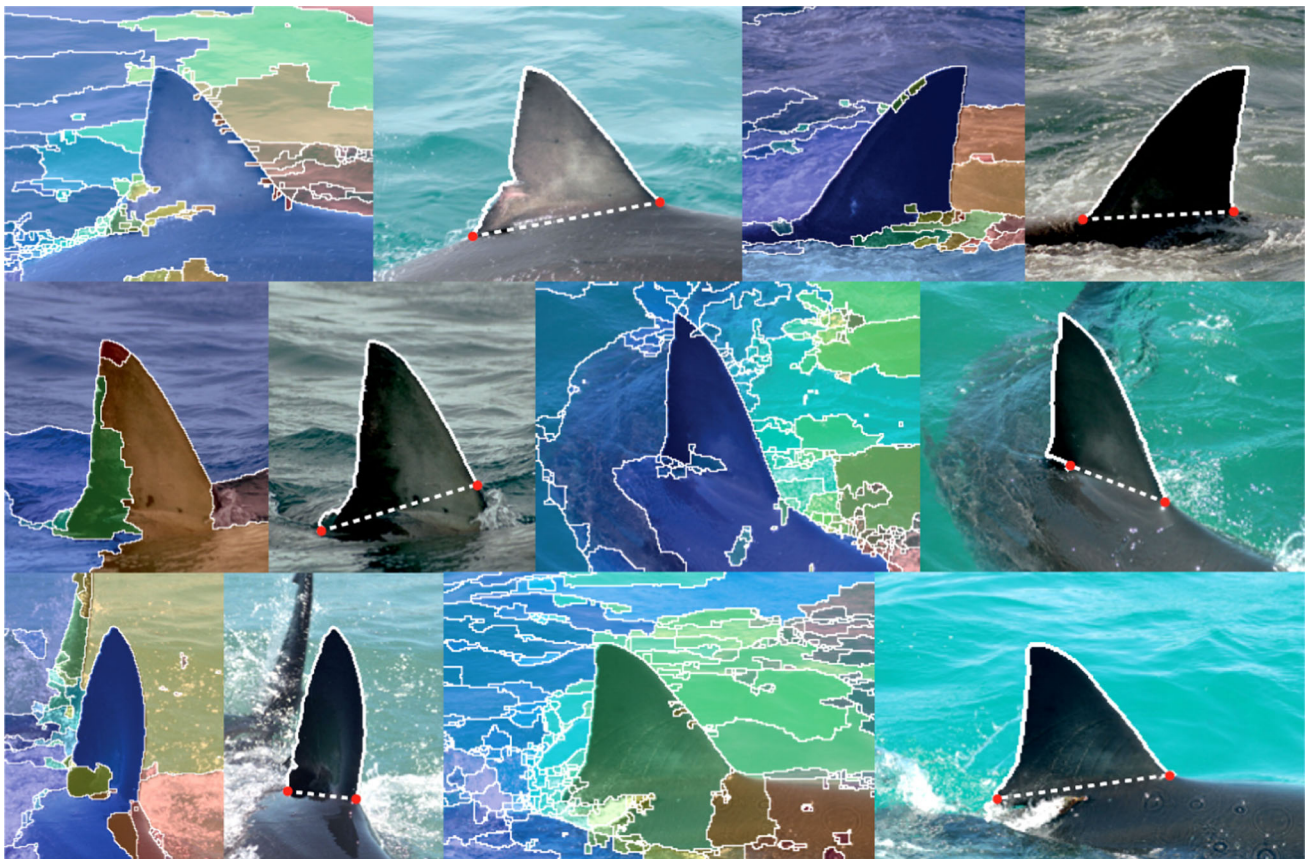
**Fig. 2** Fin detection as open contour strokes: Multi-scale 2D region-based segmentation algorithms Arbeláez et al. (2014) on their own (*left* images show one level of the ultrametric map) regularly fail to detect the extent of fins due to visual ambiguities produced by shark body, water reflections or white splash. Thus, often no level of the underly-ing 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*)
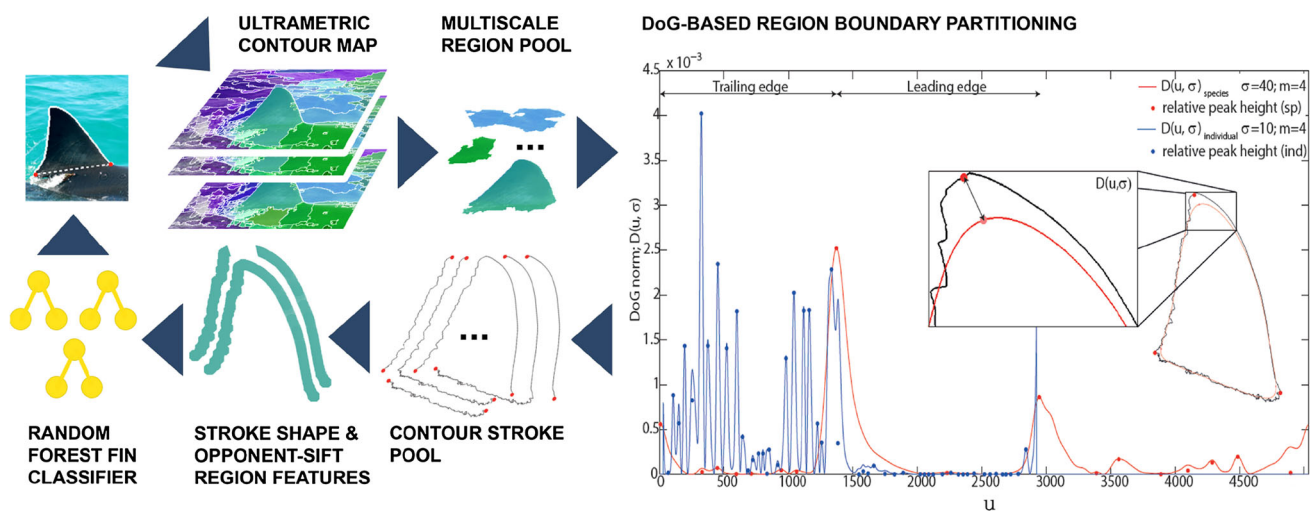


**Fig. 3** Fin detection model: partitioning the (*closed*) 2D region structures from across *all* levels of the ultrametric contour map via DoG-generated keypoints (rightmost visualisation) yields a pool of (*open*) contour strokes, whose normal-encoded shape and nearby opponentSIFT descriptors feed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

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[1]. 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.

### 3.2 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 Fig. 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 $D(u, \sigma)$:

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

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 Fig. 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 Fig. 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 $N_c = (n^2 - n)k$
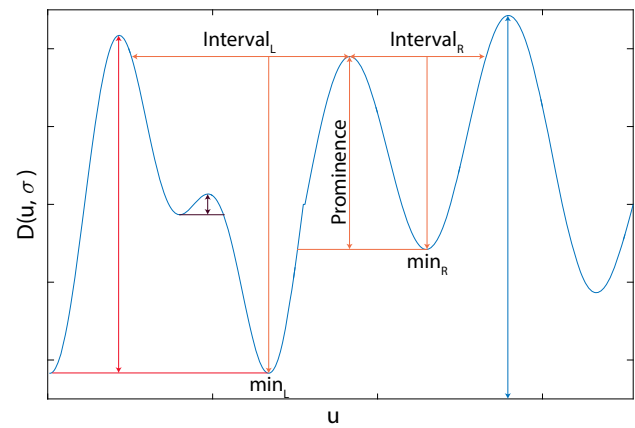
[1] Any region with a boundary length of less than 70 pixels is discarded, before the remainder are clustered into groups where all regions in a cluster have an overlap of 0.95 or more. Within each cluster, we rank regions according to the level in the hierarchy at which they first appeared, retaining the top ranked region in each cluster.

**Fig. 4** Non-maximum suppression: we utilise the Matlab function 'findpeaks' as a reference implementation for non-maximum suppression. That is, from a local maximum on $D(u, \sigma)$, the *horizontal* distance to $D(u, \sigma)$ is measured to define *left* and *right intervals* $\min_L = \min_{\text{interval}_L}(D(u, \sigma))$, $\min_R$ is defined likewise. Subsequently, $\max(\min_L, \min_R)$ is taken as a reference level. The prominence of each local maximum is then computed as the difference between the value of $D(u, \sigma)$ at the local maximum and the reference level. Low prominence peaks are suppressed. If either interval reaches the end of the signal, we set its minimum to be zero

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 $F_{\text{inst}}^g$. Evaluated with respect to a human-labelled ground truth contour, we use the standard $F$-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 parametrisation 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 $F_{\text{inst}}^g = 0.97$ 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 $F_{\text{inst}}^g = 0.75$.

### 3.3 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 enables us to establish whether the trained model is representative given partial occlusion. Examples of each image type are shown in Figures 5.

**Fig. 5** High and lower visibility fin images: the *top* row shows examples of lower visibility fin images where parts of the fin are occluded by *water line* and *white splash*. The *bottom* row shows high visibility fin images—the entire extent of the fin is visible

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 (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 $F_{\text{inst}}^{p}$. Figure 6 illustrates example candidates together with their scores.



**Fig. 6** Example fin candidates and predicted quality ($F_{\text{inst}}^{p}$). (*Top*) candidates and their scores after non-maximum suppression. (*Other*) Candidates and scores from region around the fin before non-maximum suppression. The predictive ability of the model is reflected in the stroke quality predictions for strokes describing at least part of the fin. It is unsurprising that the model makes high quality-predictions for the caudal fin stroke. We also see that while higher scores are sometimes predicted for purely background objects, the scores predicted for these are typically not as high as those predicted for good quality strokes describing fins themselves

---

[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.
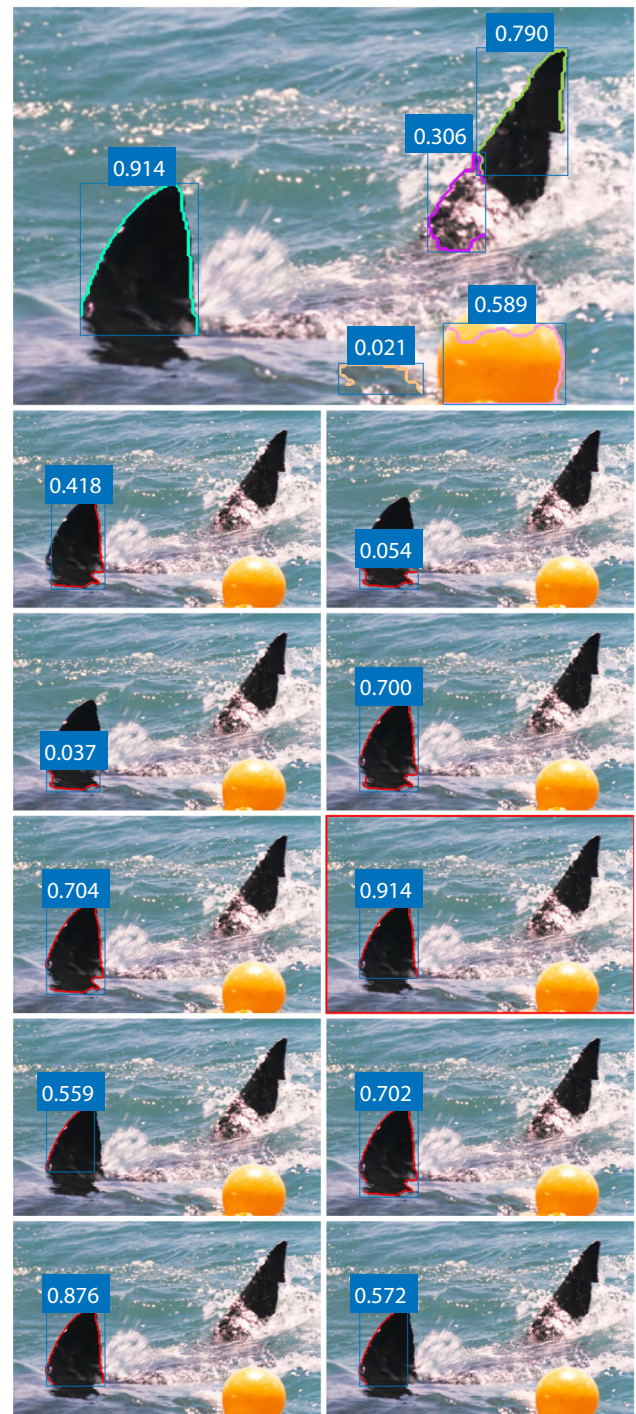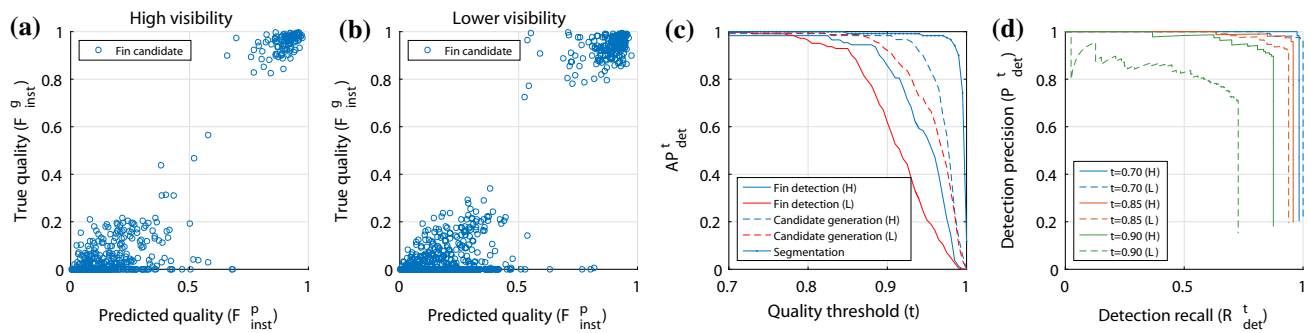
**Fig. 7** Fin detection results: **a**, **b** scatter plots show that the full fin detection model is able strongly to predict, as captured by $F_{inst}^p$, the true quality of fin candidates $F_{inst}^g$ for both high and low visibility images. **c** The plot summarises performance at different stages of fin detection.

Note, that for the 'segmentation' line, $AP_{det}^t$ is equivalent to the proportion of fins for which it is possible to obtain a stroke of quality $F_{inst}^g \geq t$, given a machine generated segmentation. **d** The plot shows PR plots for both high and low visibility images at different thresholds

## 3.4 Measuring Detection Performance

We use (1) average precision ($AP_{det}^t$), the area under the precision-recall (PR) curve for a given threshold $t$, and (2) volume under PR surface ($AP^{vol}$) as evaluation metrics.

In order to generalise $AP_{det}^t$, the $AP^{vol}$ measure was proposed by Hariharan et al. (2014) for simultaneous object detection and segmentation. It measures the volume under a PR surface traced out by PR curves generated for variable quality thresholds $t$, and thus avoids arbitrary threshold choices. It reflects both fin detection performance and the quality of candidates detected and, as noted by Hariharan et al. (2014), has the attractive property that a value of 1 indicates perfect candidate generation as well as fin detection.

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.

## 3.5 Fin Detection Results

Results for fin candidate generation and detection are shown in Fig. 7, Tables 1 and 2. Scatter plots in Fig. 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 $F_{inst}^g$ scores. The Pearson correlation coefficients

**Table 1** Intermediate results ($AP_{det}^t$)

|  | t = 0.7 | t = 0.85 | t = 0.9 | $AP^{vol}$ |
|---|---|---|---|---|
| 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 |

between true and predicted quality scores are 0.95 and 0.93, respectively.

The plot of Fig. 7(C) summarises performance at different stages of fin detection. We note that for segmentation, a stroke of quality $F_{inst}^g \geq 0.95$ is possible for almost all fin instances (98.3 %), with an average achievable quality, $AP^{vol}$, of 0.99. Candidate generation also performs well. It can be seen that for almost all high visibility fins (98.3 %), a candidate of $F_{inst}^g > 0.9$ is generated and $F_{inst}^g > 0.85$ 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 Fig. 7(C) and Table 2, that of combined segmentation, candidate generation and candidate classification. Here we see that a candidate of quality $F_{inst}^g \geq 0.83$ is generated and recognised (with $AP_{det}^t = 0.98$) for almost all high visibility fins ($F_{inst}^g \geq 0.78$ for lower visibility with $AP_{det}^t = 0.99$), as indicated by $AP_{det}^t$ close to 1 for these quality thresholds, with $AP_{det}^t = 1$ only possible if both $P_{det}^t = 1$ and $R_{det}^t = 1$.

To fully understand values of $AP_{det}^t < 1$, we must consider detection precision and recall separately, as shown in Fig. 7(D). Here we show PR curves for selected quality thresholds of the complete detection pipeline. We see for example that for $t = 0.85$, perfect precision ($P_{det}^t = 1.0$) is achieved for about 63 % of both high and lower visibility fins ($R_{det}^t = 0.63$), after which, false positives are introduced as shown by reduced precision. We also see that detection recall

**Table 2** Fin detection results ($AP_{det}^t$)

| Feature type | t = 0.7 | t = 0.85 | t = 0.9 | $AP^{vol}$ |
|---|---|---|---|---|
| High visibility (H) | | | | |
|   OpponentSIFT | 0.99 | 0.85 | 0.73 | - |
|   Normal | 0.98 | 0.85 | 0.7 | - |
|   Combined | 0.98 | 0.95 | 0.86 | 0.92 |
| Lower visibility (L) | | | | |
|   Combined | 1.0 | 0.93 | 0.62 | 0.89 |

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 $F_{inst}^g \geq 0.7$.

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 nuisance 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.

## 4.1 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 Kambhamettu (2009). This produces an opacity mask $\alpha$ 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 as a precursor to biometric encoding. Full details of this edge refinement procedure can be found in Hughes and Burghardt (2015a).

## 4.2 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 Sect. 3. To generate boundary subsections, we resample fin candidates to a fixed resolution of 1024 pixels and compute $D(u, \sigma)$ in Eq. 1, re-parametrised with $\sigma = 2$ and $m = 8$. Subdivision by these keypoints yields $\binom{50}{2} = 1225$ 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 Fig. 8.

## 4.3 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 Eq. 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 Fig. 10).

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 concate-

---

[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.
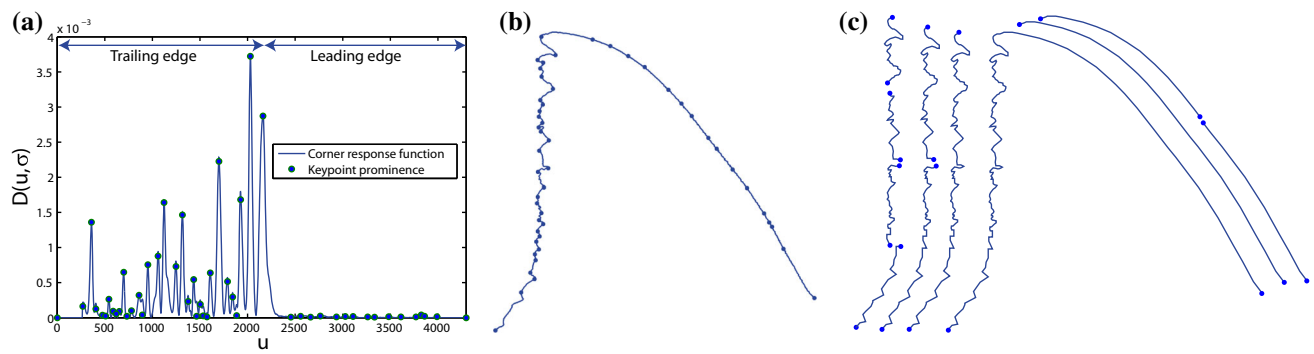
**Fig. 8** Combinatorial contour sampling: **a** the DoG corner response function of a fin contour. **b** The $n = 50$ most prominent maxima of $D$ are selected as keypoints. The detected keypoints are shown on the fin contour. **c** The contour is combinatorially sampled between every keypoint pair to produce a set of local, semi-local and global subsections

nated to yield the descriptor:

$$\mathcal{N}(u, \sigma) = (G(u, \sigma) * x(u), G(u, \sigma) * y(u)) \qquad (2)$$

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.

As illustrated in Fig. 9 over the entire fin segment, both $\text{DoG}_N$ 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.

Specifically, denoting the set of descriptors for a query object $D_Q$, for each query descriptor $d_i \in D_Q$ and for each class $c \in C$, we find two nearest neighbours $(NN_c(d_i), NN_{\bar{C}}(d_i))$ where $\bar{C}$ is the set of all classes other than $c$. Using the shorthand $\delta(NN.) = ||d_i - NN.(d_i)||^2$, queries are classified according to:

$$\hat{C} = \arg\max_C \sum_{i=1}^{|D_Q|} f(d_i, c) \qquad (3)$$

$$f(d, c) = \begin{cases} \delta(NN_{\bar{C}}) - \delta(NN_c) & \delta(NN_{\bar{C}}) > \delta(NN_c) \\ 0 & \text{otherwise} \end{cases} \qquad (4)$$

This decision rule can be extended to a multi-scale case. Letting $S = \{\sigma_1, \ldots, \sigma_j, \ldots, \sigma_v\}$ denote the set of scales for which we compute descriptors, the multi-scale decision rule
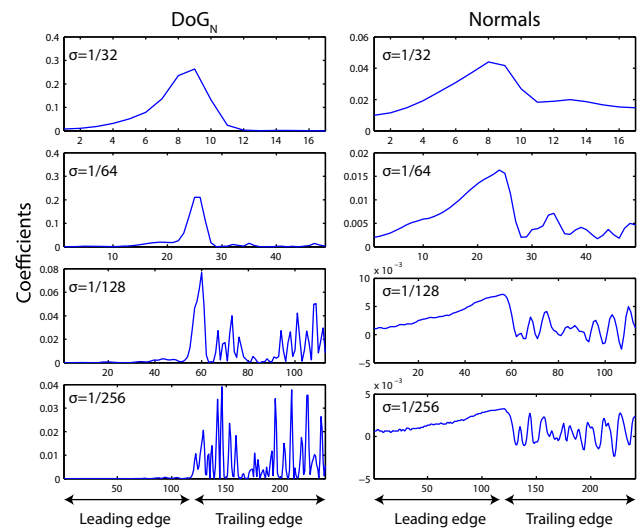


**Fig. 9** Descriptors for encoding individual fin shape: we utilise the $\text{DoG}_N$ and Arandjelovic's normal descriptor as a feature pool for characterising individuality. It can be seen that both location on the segment ($x$-axis) and scale-space band ($\sigma$) are encoded by the descriptors

linearly combines the contribution of the descriptors at each scale (see also top of Fig. 15):

$$\hat{C} = \arg\max_C \sum_{j=1}^{v} w_j \cdot \sum_{i=1}^{|D_Q^j|} f(d_i^j, c) \qquad (5)$$

### 5.1 Implementation Details

To achieve scale normalisation, each contour subsection is re-sampled to a fixed length of 256 pixels. $\text{DoG}_N$ and normal descriptors are computed at filter scales $S = \{1, 2, 4, 8\}$, with a constant value of $m = 2$ in the $\text{DoG}_N$ 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 ($w_j = 1$).

### 5.2 Dataset

In order to benchmark individual fin classification, we use a dataset representing 85 individuals and consisting of 2456 images (see Acknowledgements 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 Figs. 1, 10 for example images).

### 5.3 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 Eqs. 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).

### 5.4 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 0.79. Figure 10 illustrates such examples of fin matches. An examination of recognition performance for high quality fin detections ($F_{inst}^g > 0.9$) provides insight into the effect 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 could not be clas-

**Table 3** Individual LNBNN ID Results

| Encoding | $\sigma = 8$ | $\sigma = 4$ | $\sigma = 2$ | $\sigma = 1$ | Combined |
|---|---|---|---|---|---|
| 1 Training image per class (1-shot-learning): 2371 queries | | | | | |
| 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 | | | | | |
| AP:SIFT | | | | | 0.20 |
| mAP:SIFT | | | | | 0.35 |
| AP:DoG$_N$ | | | | | 0.81 |
| mAP:DoG$_N$ | | | | | 0.83 |

sified 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.

Figure 11 shows precision-recall curves for DoG$_N$ and normal encoding types. It can be seen that the recognition performance difference between the two feature types occurs in the high precision region, with a normal encoding providing recognition precision of less than one for almost all values of recall. When descriptors corresponding to the trailing edge of fins alone are considered, the normal encoding provides superior recognition to that obtained using DoG$_N$, but nevertheless remains inferior to that obtained using a multi-scale DoG$_N$ representation of the whole fin.

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 Sect. 6.

### 5.5 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
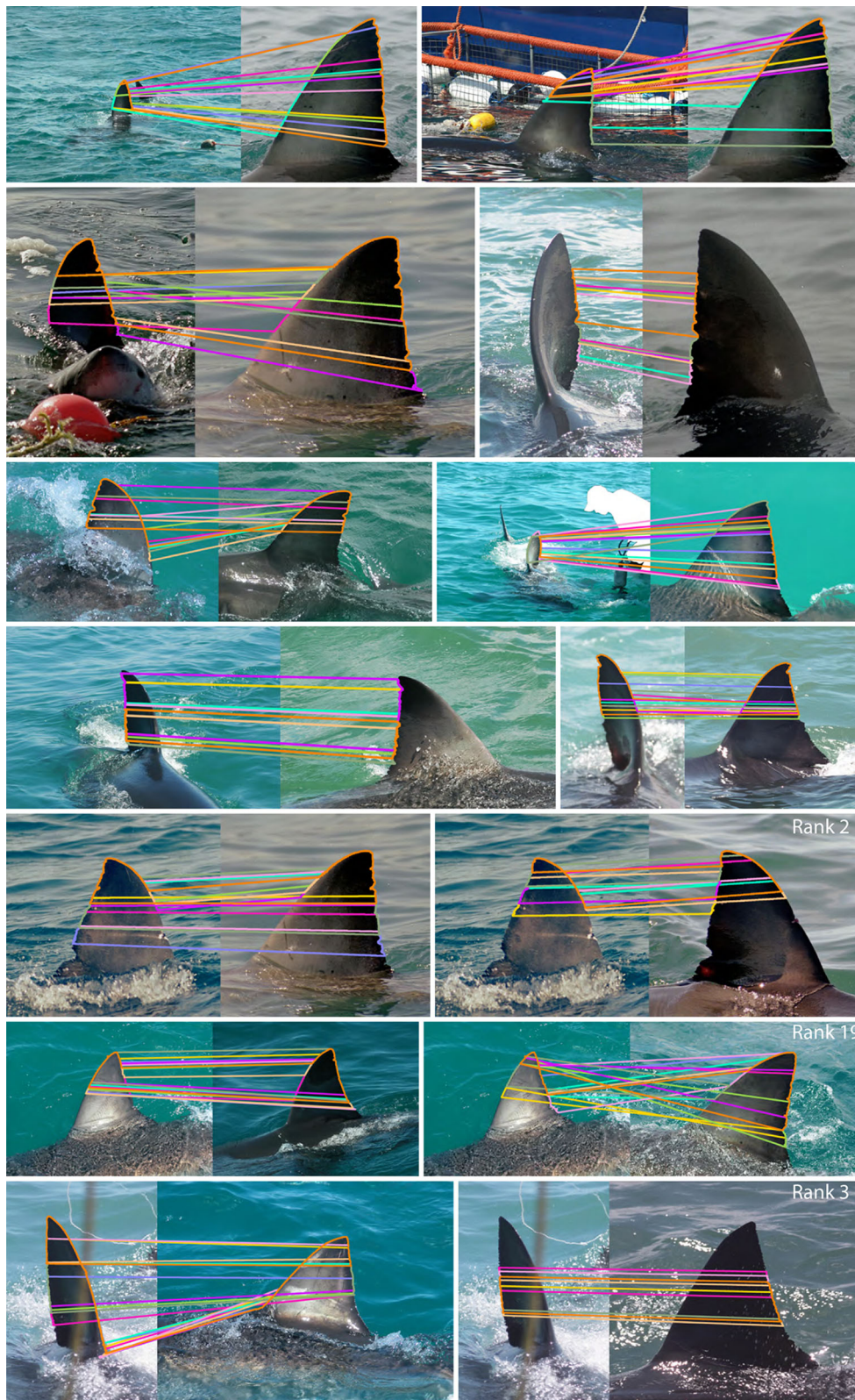
**Fig. 10** LNBNN individual identification examples: *left* images are queries and right ones are predicted individuals. *Coloured lines* indicate start and end of the ten sections contributing most evidence for the matched individual. For illustration of false matches, *bottom* three rows, left pairs, show misidentifications while correct matches are shown right. All example matches are obtained using multiscale DoG$_N$ 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
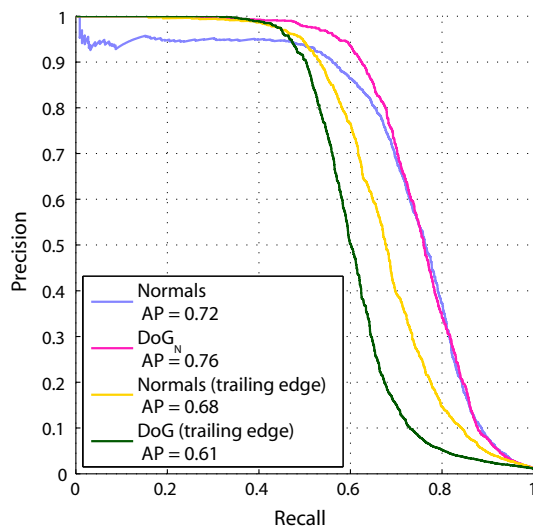
**Fig. 11** Precision-recall curves for LNBNN. Precision-recall curves for $DoG_N$ and normal fin encodings, comparing identification via whole fins and just trailing edges
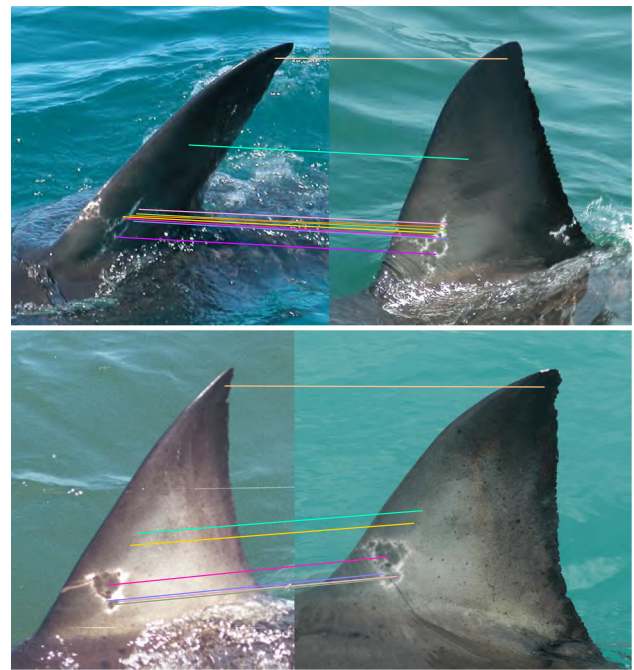


**Fig. 12** Example identifications using affine-covariant sift descriptions: rarely, fins misclassified using biometric contour representations are correctly identified using surface texture descriptors. Here, two such examples are shown, with query images on the left of each pair. The *coloured lines* represent discriminative feature matches (as evaluated by the value of $f(d, c)$ in Eq. 4)

fin stokes (illustrated using dashed lines in Fig. 2). As before, LNBNN (Eqs. 3, 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 $DoG_N$ encoding (AP = 0.81). Interestingly however, 45 fin instances, misclassified using biometric contour encoding, are correctly identified using SIFT, with examples shown in Fig. 12. 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 $DoG_N$ 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 4$D$ fin space is illustrated in Fig. 13.

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.

## 6.1 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 Fig. 14, 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

---

[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.
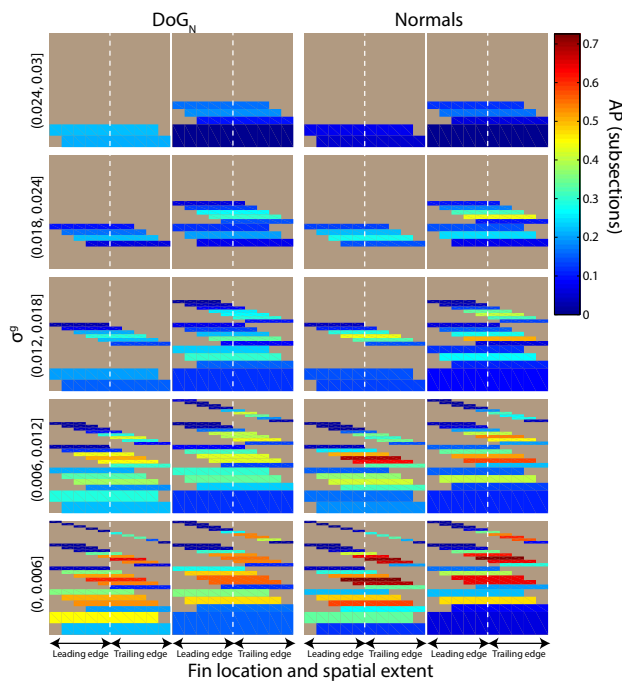
**Fig. 13** Fin space and localisation of individuality. Organising visual descriptors indexed over spatial location (*x*-axes) and extent on the fin (*dotted line marks* fin tip), and filter scale (*rows*) allows for the learning of population-wide distinctiveness properties associated with the anatomic fin locations. Colouration depicts the distinctiveness of bins with respect to animal individuality, as quantified by classification AP at the subsection level
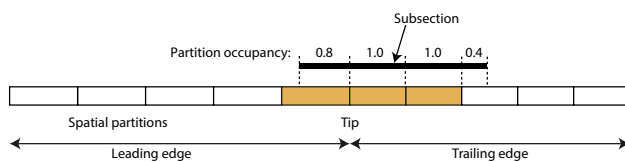


**Fig. 14** Spatial embedding of fin patterns. Example of a subsection mapped to a spatial bin (shown in *yellow*) covering 3 partitions (Color figure online)

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 $B = \{(\sigma_1^g, \sigma_2^g], \ldots, (\sigma_k^g, \sigma_{k+1}^g], \ldots, (\sigma_{m-1}^g, \sigma_m^g]\}$ and the set of filter scales is $S^g = \{\sigma_1^g, \ldots, \sigma_k^g, \ldots, \sigma_m^g\}$. 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 Eq. 5) can be expressed as $\sigma_{i,j}^g = \sigma_j / l_n \cdot p$ where $p$ expresses the length of the subsection as a proportion of the length of the fin contour, and $l_n$ is the number of samples used to encode 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.

## 7.1 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 Sect. 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 Eq. 4.

Now, instead of sum-pooling local match scores over class labels directly, as performed in Eqs. 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 Fig. 15). 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.

## 7.2 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 Sect. 5), exploit the observed variance in local distinctiveness (see Fig. 13), 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.
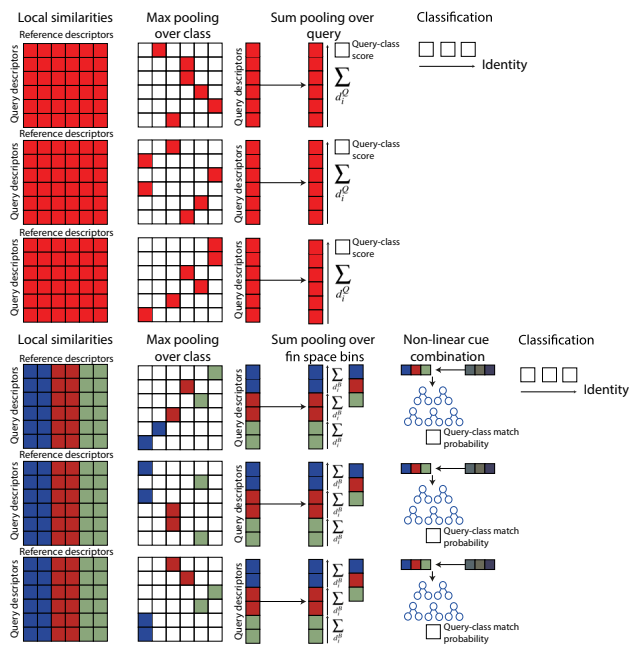
**Fig. 15** Comparison of baseline (*top*) and fin-space identification scheme (*bottom*). The two paradigms are illustrated proceeding from *left* to *right*. By associating descriptor matching scores (*left column*) to reference locations in a global fin space (*colouration*), the improved scheme (*bottom*) 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

### 7.3 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 Sect. 5. We present the results in Fig. 16. 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.

## 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.
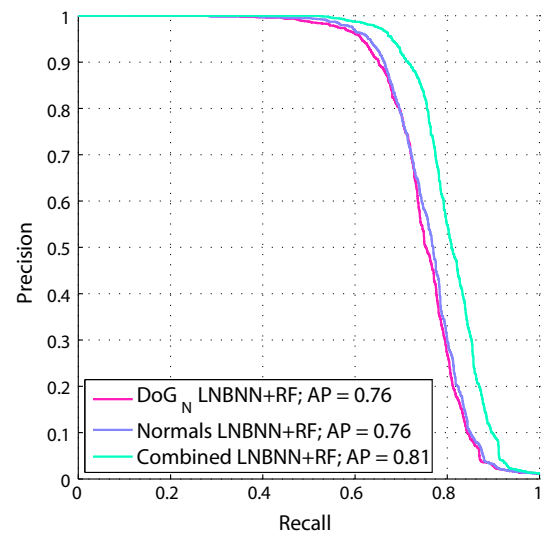


**Fig. 16** Results of identification using the fin space approach. Precision-recall curves reported considering each of the descriptor types separately (effectively training the random forest on only half the fin space), as well as considering the full dual descriptor set

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

## References

Anderson, S. D., Chapple, T. K., Jorgensen, S. J., Klimley, A. P., & Block, B. A. (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* (pp 1–11). Guildford: BMVA Press.

Arandjelovic, R., & Zisserman, A. (2011). Smooth object retrieval using a bag of boundaries. In *2011 IEEE international conference on computer vision (ICCV)* (pp 375–382). IEEE.

Arbeláez, P., Hariharan, B., Gu, C., Gupta, S., Bourdev, L., & Malik, J. (2012). Semantic segmentation using regions and parts. In *2012 IEEE conference on computer vision and pattern recognition (CVPR)* (pp 3378–3385). IEEE.

Arbeláez, P., Pont-Tuset, J., Barron, J. T., Marques, F., & Malik, J. (2014). Multiscale combinatorial grouping. In *CVPR*.

Baird, R. W., Gorgone, A. M., McSweeney, D. J., Webster, D. L., Salden, D. R., Deakos, M. H., et al. (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, R. W., Gorgone, A. M., McSweeney, D. J., Ligon, A. D., Deakos, M. H., Webster, D. L., et al. (2009). 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, T. Y., Rubenstein, D. I., Stewart, C. V., 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 *CVPR 2008. IEEE conference on computer vision and pattern recognition, 2008* (pp 1–8). IEEE.

Bonfil, R., Meÿer, M., Scholl, M. C., Johnson, R., O'Brien, S., Oosthuizen, H., et al. (2005). Transoceanic migration, spatial dynamics, and population linkages of white sharks. *Science*, *310*(5745), 100–103.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Carreira, J., & Sminchisescu, C. (2010). Constrained parametric min-cuts for automatic object segmentation. In *2010 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 3241–3248). IEEE.

Chapple, T. K., Jorgensen, S. J., Anderson, S. D., Kanive, P. E., Klimley, A. P., Botsford, L. W., et al. (2011). A first estimate of white shark, carcharodon carcharias, abundance off Central California. *Biology Letters*, *7*(4), 581–583.

Crall, J. P., Stewart, C. V., Berger-Wolf, T. Y., Rubenstein, D. I., & Sundaresan, S. R. (2013). Hotspotterpatterned species instance recognition. In *2013 IEEE workshop on applications of computer vision (WACV)* (pp. 230–237). IEEE.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *CVPR 2005. IEEE computer society conference on computer vision and pattern recognition, 2005* (Vol. 1, pp. 886–893). IEEE.

Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and roc curves. In *Proceedings of the 23rd international conference on machine learning* (pp. 233–240). ACM.

Dollár, P., & Zitnick, C. L. (2013). Structured forests for fast edge detection. In *2013 IEEE international conference on computer vision (ICCV)* (pp. 1841–1848). IEEE.

Duyck, J., Finn, C., Hutcheonb, A., Vera, P., Salas, J., & Ravela, S. (2015). Sloop: A pattern retrieval engine for individual animal identification. *Pattern Recognition*, *48*, 10591073.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, *88*(2), 303–338.

Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2014). The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, *111*(1), 98–136.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2010). Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(9), 1627–1645.

Freytag, A., Rodner, E., Simon, M., Loos, A., Khl, H., & Denzler, J. (2016). Chimpanzee faces in the wild: Log-euclidean CNNs for predicting identities and attributes of primates. In *German conference of pattern recognition (GCPR)*.

Geusebroek, J. M., Burghouts, G. J., & Smeulders, A. W. (2005). The amsterdam library of object images. *International Journal of Computer Vision*, *61*(1), 103–112.

Gu, C., Lim, J. J., Arbeláez, P., & Malik, J. (2009). Recognition using regions. In *CVPR 2009. IEEE conference on computer vision and pattern recognition, 2009* (pp. 1030–1037). IEEE.

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

Hillman, G., Gailey, G., Kehtarnavaz, N., Drobyshevsky, A., Araabi, B., Tagare, H., et al. (2003). Computer-assisted photo-identification of individual marine vertebrates: A multi-species system. *Aquatic Mammals*, *29*(1), 117–123.

Hughes, B., & Burghardt, T. (2015a). Affinity matting for pixel-accurate fin shape recovery from great white shark imagery. In *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. In *24th British machine vision conference, (BMVC)* (pp. 92.1–92.14).

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

Kühl, H. S., & 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 *2010 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 1712–1719). IEEE.

Loos, A., & Ernst, A. (2013). An automated chimpanzee identification system using face detection and recognition. *EURASIP Journal on Image and Video Processing*, *2013*, 1–17. 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, D. R., Fowlkes, C. C., & Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*(5), 530–549.

McCann, S., & Lowe, D. G. (2012). Local naive bayes nearest neighbor for image classification. In *2012 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 3650–3656). IEEE.

Mikolajczyk, K., & Schmid, C. (2004). Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, *60*(1), 63–86.

Muja, M., & Lowe, D. G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. *VISAPP*, *1*, 2.

Ranguelova, E., Huiskes, M., & Pauwels, E. J. (2004). Towards computer-assisted photo-identification of humpback whales. In *2004 International conference on image processing, 2004. ICIP'04* (Vol. 3, pp. 1727–1730). IEEE.

Ravela, S., Duyck, J., & Finn, C. (2013). Vision-based biometrics for conservation. In Carrasco-Ochoa, J. A., Martnez-Trinidad, J. F., Rodrguez, J. S., & di Baja., G. S. (Eds.), *MCPR 2012 LNCS 7914* (pp. 10–19). Berlin: Springer.

Robbins, R., & Fox, A. (2013). Further evidence of pigmentation change in white sharks, carcharodon carcharias. *Marine and Freshwater Research*, *63*(12), 1215–1217.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., 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, K. R., & Roure, D. D. (2013). Zooniverse: Observing the world's largest citizen science platform. In *Proceedings of the 23rd international conference on world wide web* (pp. 1049–1054).

Speed, C. W., Meekan, M. G., & Bradshaw, C. J. (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., Debure, 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* (pp. 648–660). New York: Springer.

Towner, A. V., Wcisel, M. A., Reisinger, R. R., Edwards, D., & Jewell, O. J. (2013). Gauging the threat: The first population estimate for white sharks in South Africa using photo identification and automated software. *PloS One*, *8*(6), e66035.

Uijlings, J. R., van de Sande, K. E., Gevers, T., & Smeulders, A. W. (2013). Selective search for object recognition. *International Journal of Computer Vision*, *104*(2), 154–171.

Van De Sande, K. E., Gevers, T., & Snoek, C. G. (2010). Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(9), 1582–1596.

Van Hoey, N. E. (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 raggedtooth 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 *CVPR 2001. Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition, 2001* (Vol. 1, pp I–511). IEEE.

Zhang, X., Wang, H., Hong, M., Xu, L., Yang, D., & Lovell, B. C. (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 *2009 IEEE 12th international conference on computer vision* (pp. 889–896). IEEE.