Automated Identification of Individual Great White Sharks from Unrestricted Fin Imagery

Benjamin Hughes
benjamin.hughes@bristol.ac.uk

Tilo Burghardt
tilo@cs.bris.ac.uk

Department of Computer Science
University of Bristol
Bristol, UK

Introduction. Recognising individuals repeatedly over time is a basic requirement for field-based ecology and related life sciences [5, 6]. In this paper we propose a visual identification approach for great white shark fins as outlined in Figure 1, one that is applicable to unconstrained fin imagery and fully automates the pipeline from feature extraction to matching of identities. We pose the associated vision task as a fine-grained, multi-instance classification problem for flexible, smooth and partly occluded object parts. To the best of our knowledge, the proposed system is the first fully automated contour-based animal biometrics system.

Contour Stroke Object Model. A key technical contribution of the paper is a contour stroke model trained for fin detection. It combines a partitioning of ultrametric contour maps (UCM; generated via [3, 4]) with normal descriptors, dense local features and a random-forest regressor. UCM region boundaries are partitioned at keypoints detected as local maxima of the Difference of Gaussian (DoG) function \( D(u, \sigma) \) detailed by Zhang et al. [10]. We generate fin candidates as contour strokes by sampling the region contour between every permutation of keypoint pairs. Each stroke is then described by a 160 dimensional feature vector consisting of two components: the first is a bag of opponent SIFT [9] visual words, the second describes boundary shape using a histogram of boundary normals. A random forest regressor is finally trained to predict the fin-like quality of contour stroke candidates. Figure 2 illustrates fins detected by this strategy. In common with the segmentation stage of Arandjelovic and Zisserman’s ‘Bag of Boundaries’ pipeline [2], our contour stroke model separates objects in natural images and against cluttered backgrounds, but differs in that it exploits species characteristic shape encoded in the open contours, to specifically and explicitly detect fin parts.

Biometric Encoding: [1, 2] perform smooth object recognition using semi-local boundary normal and HoG-occupancy representations, respectively. These allow efficient and robust matching, but their encoding of inter-class variance will always be sub-maximal. Meanwhile, the semi-automated fin recognition system DARWIN [8] uses a global 2D Cartesian representation. This encoding maximises inter-class variance, but is sensitive to partial occlusions and detection errors. By contrast we utilise semi-local and global shape descriptions in a vector-based combinatorial encoding strategy, one that enables efficient and accurate individual recognition while being robust to noisy, partially occluded input.

As with generating contour strokes, we combinatorially sample boundary sections between pairs of keypoints detected using \( D(u, \sigma) \), reparameterised for fine-grained feature detection. We then scale-normalise the resulting contour subsections and encode their shape over various spectral scales. Denoting the set of descriptors for a query object at scale \( j \) as \( D_Q^j \), for each query descriptor \( d_j^i \in D_Q^j \), we find the two nearest neighbours \((NNC(d_j^i), NNC(d_j^{i+1}))\) where \( C \) is the class of the nearest neighbour and \( C \) is the set of all other classes. Subsection information is then combined using a multi-scale decision rule based on the Local Naive Bayes Nearest Neighbour (LNBNN) classifier [7]:

\[
\hat{C} = \arg \max_C \sum_j w_j \sum_i | |d_j^i - NNC(d_j^i)|^2 - | |d_j^{i+1} - NNC(d_j^{i+1})|^2 |
\]

Results: We benchmark system performance using a dataset presenting 85 individuals and consisting of 2456 images (one labelled reference image per shark), exhibiting significant variability in waterline and splash occlusion, viewpoint, orientation and scale (see various Figures). Overall, 72% of queries are correctly identified with a mAP of 0.79. Discounting detection errors and partial occlusions, this rises to 82% accuracy with a mAP of 0.84. Fin detection performance depends on the quality threshold \( t \) for detected fin candidates, as measured by \( F \) - scores. For unoccluded instances we obtain: \( AP=0.98 \) for \( t > 0.7 \) and \( AP=0.85 \) for \( t > 0.9 \).

Conclusion: We have combined a contour stroke model with a spectral and combinatorial contour encoder. The method shows promising individual shark fin identification performance. We conclude applicability at accuracy levels ready to assist human identification without labelling.