Affinity Matting for Pixel-accurate Fin Shape Recovery from Great White Shark Imagery

Benjamin Hughes
benjamin.hughes@bristol.ac.uk
Tilo Burghardt
tilo@cs.bris.ac.uk

Department of Computer Science
University of Bristol
Bristol, UK

Abstract

The objective of this paper is to obtain pixel-accurate reconstructions of white shark fins given automatically generated coarse pre-segmentations. Reconstruction performance is compared for affinity matting, colour matting and GrabCut against expert annotated ground truth for a test-set of 120 fin images taken in the wild. For the present domain, we find affinity matting able to most accurately recover fine shape details, whilst being robust to wide baseline trimap initialisations as needed to reconstruct prominent notches on the fin edge.

1 Introduction

Re-recognising individual animals over time is a fundamental task in field-based biology [10]. Whenever animals carry individually unique visual markings and an approach for imaging these efficiently is available, biometric computer vision provides an option for non-invasive, partly or fully automated identification [9, 12, 14, 15]. Individual great white sharks, for instance, can be fully automatically re-identified if one can photograph and match silhouettes of their dorsal fin [7], but this requires a precise (and fully automatic) extraction of boundaries as a precursor to matching. State-of-the-art image segmentation frameworks [1, 2] may be used as locally coarse boundary detectors, but they rarely provide the levels of fine grained segmentation accuracy required for identification, particularly when applied at image resolutions optimised for efficiency[1]. On the other hand, contour segmentation methods [8, 11] applied in existing semi-automated approaches to fin recognition [4, 5, 14] can achieve accurate and *computationally* efficient segmentation results, but rely on user interaction for initialisation and to correct segmentation errors.

In this paper we investigate affinity matting, as described in [17], for reconstructing fin shape details for great white sharks given automatically generated, locally coarse pre-segmentations. We evaluate segmentation accuracy against colour matting [17] and GrabCut [13] on a pixel-accurately annotated dataset of 120 fin images of great whites photographed in the wild (see Figure 1).

2 Affinity-based Matting

Building on the given coarse fin boundary $\hat{C}$, our task is to generate a refined boundary $\hat{C}'$ which matches closely the actual fin shape as labelled by the ground truth. To do this we

[1] In [7] images are resized to 0.05MP prior to segmentation from an average full image resolution of 4MP. Following object detection, fin boundaries are resized to their original scale through a process of rescaling and interpolation, but this cannot recover fine details lost as a result of downsampling.

Pages 8.1-8.8
https://dx.doi.org/10.5244/C.29.MVAB.8

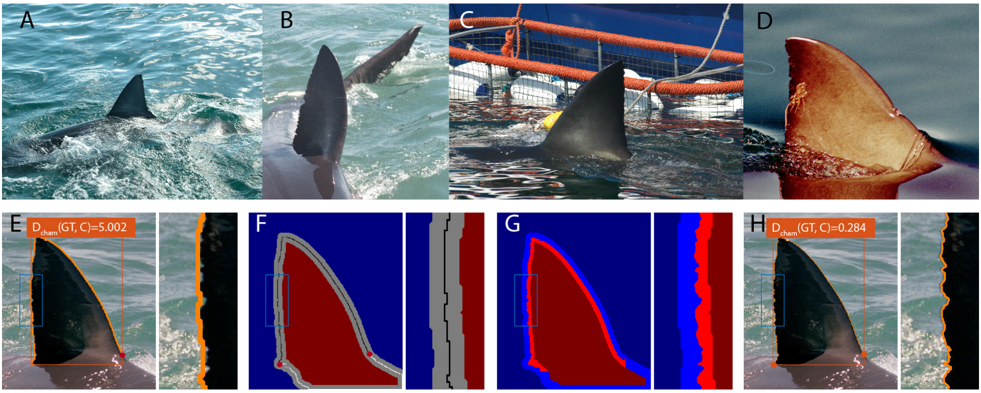

Figure 1: **Dataset and Task. (A-D)** Samples of fin images of great white sharks taken under natural conditions. We use a dataset of 120 fin images, representing 50 individuals. The average height of fins, measured by the distance between the tip and the midpoint of the line segment connecting the two ends of the fin contour, is 970 pixels (standard deviation 327 pixels). Given a coarse fin boundary $\hat{C}$ **(E)** and an associated trimap **(F)**, this work studies matting techniques for achieving pixel-accurate fin segmentations **(G,H)**. Chamfer distances to the ground truth *GT* before and after affinity matting are also shown.

first subdivide the image into a trimap containing 'definitely fin', 'definitely background' and 'potential boundary' regions (see Figure 1). Following notations in [17], for an image I, we formally denote the set of all n pixels as $\Omega = \{1, ..., i, ..., n\}$. Labelled pixels for definite fin region Ω_l^f and background Ω_l^b then combine to all labelled pixels $\Omega_l \equiv \Omega_l^b \cup \Omega_l^f$. This is complemented by the set of unlabelled 'potential pixels' $\Omega_u = \Omega \backslash \Omega_l$, that is pixels within a distance D_u of the given approximate fin boundary. Under the notion of 'matting', the image I is interpreted as a linear combination of fin F and background B images (see Figure 2) mixed via a blending mask α:

$$I = \alpha F + (1 - \alpha) B \tag{1}$$

where $\alpha_i \in [0.1]$. Thus, at initialisation we have a set of labelled pixels Ω_l, with known alpha values 0 or 1, and a set of unlabelled pixels Ω_u for which we wish to calculate alpha.

Wang et al. [16] distinguish three categories of approach for estimating unknown alphas: (1) colour matting, which seeks to estimate unknown alphas directly from nearby labelled pixels, (2) affinity matting, which estimates alphas from the alpha values of neighbouring pixels, and (3) a mixed category combining both approaches. In our work we will apply two methods from [17], one from each of the first two categories.

We will first introduce the estimation of alpha directly from nearby labelled pixels, i.e. describe colour matting. Using the notation from [17], two subsets $Q_l^f \subseteq \Omega_l^f$ and $Q_l^b \subseteq \Omega_l^b$ are defined such that any pixel j within a Euclidean distance $D_j < D$ away from Ω_u is selected, that is a rim around Ω_u. Subsequently, for each unknown pixel i in Ω_u, two further subsets $Q_l^{f\prime} \subset Q_l^f$ and $Q_l^{b\prime} \subset Q_l^b$ are selected as the nearest k pixels to i in Q_l^f and Q_l^b, respectively. We now determine $D = (\gamma_d |\Omega_u|)/|\Omega| + \sqrt{2}$ with $\gamma_d = 1.2$ for a setting of $k = 80$. Having obtained $Q_l^{f\prime}$ and $Q_l^{b\prime}$ they are weighted and used to learn an alpha-colour model for the neighbourhood of pixel i using weighted ridge regression [6]. Alpha is then estimated as:

$$\alpha_i = \mathbf{f}_i^T \boldsymbol{\alpha}_i \tag{2}$$

In this case α_i is the alpha value of the unknown pixel i and $\boldsymbol{\alpha}_i$ is $\boldsymbol{\alpha}_{Q_i}$, denoting the vector of alpha values of the $t = |\mathrm{Q}_l^{f\prime} \cup Q_l^{b\prime}|$ neighbouring pixels, which in this case are those in $Q_l^{f\prime}$

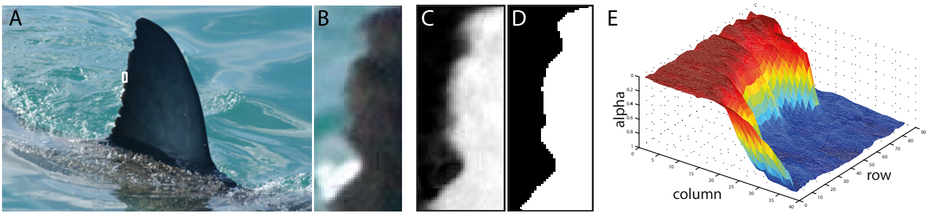

Figure 2: **Matting for Fin Shape Recovery.** **(A)** An image of a shark fin and **(B)**, a close-up part of the trailing edge. Most pixels are clearly either fully foreground or background, but inbetween pixels are mixed. **(C, E)** The alpha intensity surface. While colour information may vary considerably both across different instances of an object class or within a single instance, the alpha values are largely invariant to non-local colour change. Given the alpha channels invariance property, we classify unlabelled pixels by simply thresholding alpha at 0.5 **(D)**.

and $Q_l^{b\prime}$. The task is to learn the combination coefficients $\mathbf{f}_i$ in Equation 2, which in the colour matting case are denoted $\mathbf{f}_{Q_i}$. Each pixel j in $Q_l^{f\prime}$ and $Q_l^{b\prime}$ is set a weight $w_j = 1/(D_j)^{\gamma w}$, and from these weights, a diagonal matrix $\mathbf{W}_Q$ is constructed where the j^{th} diagonal element is w_j. The colour of any pixel is denoted as a data point $\mathbf{x} \in R^d$, where $d = 3$ are the RGB colour values of a pixel and additionally, $\mathbf{x}' = [\mathbf{x}^T 1]^T$. The colours of the neighbouring pixels, $\mathbf{x}'_j$, are encoded by building a matrix $\mathbf{X}_{Q_i}$ of size t by $(d+1)$ by stacking the labelled neighbouring data points $\mathbf{x}'_j$ such that $\mathbf{X}_{Q_i} = [\mathbf{x}'_1 ... \mathbf{x}'_t]^T$. The combination coefficients are:

$$\mathbf{f}_{Q_i} = \mathbf{x}_i'^T \mathbf{X}_{Q_i}^T \mathbf{W}_{Q_i} (\mathbf{W}_{Q_i} \mathbf{X}_{Q_i} \mathbf{X}_{Q_i}^T \mathbf{W}_{Q_i} + \lambda_r \mathbf{I}_{(t)})^{-1} \tag{3}$$

where $\mathbf{I}_{(t)}$ is the $t \times t$ identity matrix. The parameter γ^w is set to 0.25 and $\lambda_r = 0.01$.

Proceeding towards affinity matting, $\boldsymbol{\alpha}_i$ in Equation 2, which is denoted $\boldsymbol{\alpha}_{N_i}$ is now defined over a much smaller neighbourhood $N_i \subset \Omega$, which in our case we set to be a 3×3 local path centred at i. By ridge regression, the local combination coefficients of N_i are given by:

$$\mathbf{f}_{N_i} = (\mathbf{X}_{N_i} \mathbf{X}_{N_i}^T + \lambda_r \mathbf{I}_{(m)})^{-1} \mathbf{X}_{N_i} \mathbf{x}'_i \tag{4}$$

where $\mathbf{X}_{N_i}$ is constructed in the same way as $\mathbf{X}_{Q_i}$, just using RGB colour values of the m neighbours in N_i. The additional difference is that as $\boldsymbol{\alpha}_{N_i}$ cannot be measured directly, the vector $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_i, ..., \alpha n]^T; i \in \Omega$ is estimated simultaneously by minimising a quadratic cost, as described in [17].

3 Analysis of Initialisation Neighbourhood Ω_u

For applying the described matting to fin recovery, our first task is to select a suitable Ω_u as defined by D_u. Let us denote the initially given, coarse boundary to be $\hat{C}$ and the ground truth contour GT. In order to generalise about localisation errors of $\hat{C}$ and determine a workable D_u, we use contours from 120 images and two distance measures. The first is the Chamfer distance:

$$D_{cham}(GT, C) = \frac{1}{|GT|} \sum_{g \in GT} min_{c \in C} ||g - c|| \tag{5}$$

where $||.||$ denotes the Euclidean norm. This measures the mean, minimum distance from a ground truth pixel to a pixel on $\hat{C}$ and therefore measures the average localisation error induced by $\hat{C}$. In Figure 3, we plot the relationship between $D_{cham}(GT, \hat{C})$ and the length of $\hat{C}$, denoted $l_{\hat{C}}$. Note the linear relationship between mean localisation error and object scale, with the distance given by $D_{cham}(GT, \hat{C}) = 0.001 l_{\hat{C}} + 4$ pixels. Secondly, we measure the maximum localisation error of GT by $\hat{C}$ via the one-sided Hausdorff distance:

$$D_{haus}(GT,C) = max_{g \in GT} min_{c \in C} ||g - c|| \tag{6}$$

This is important to measure as we find maximum localisation errors typically occur for prominent notches on the fin edge. In Figure 3 we plot this maximum localisation error per instance against contour length, which we find to be linearly dependent on scale too. Accounting for a 9-pixel constant error by subtracting it from D_{haus}, we see that the Hausdorff distance for any ground truth instance is 0.0155 $l_{\hat{C}}$, while the average maximum is estimated to be 0.0040 $l_{\hat{C}}$. These observations suggest an optimal value of D_u in the range $[0.001 l_{\hat{C}} + 4, 0.0155 l_{\hat{C}} + 9]$.

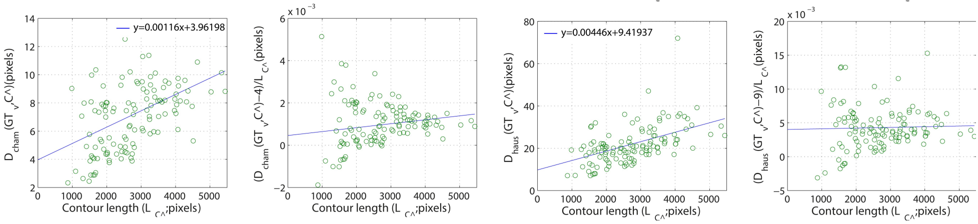

Figure 3: **Initial Localisation Error. (First)** Average localisation errors from ground truth to initial coarse boundaries. Green circles represent fin instances. The localisation error is linearly dependent on object scale, plus a scale-invariant error of around 4 pixels. **(Second)** Relative localisation errors. **(Third)** Maximum localisation errors per fin instance. **(Fourth)** The normalised Hausdorff-based error is scale-invariant, that is relative to candidate length and adjusted for a constant error of 9 pixels.

4 Contour Reconstruction Results

In order to benchmark the reconstruction accuracy we can achieve via matting, we now compare obtained fin reconstructions, denoted $\hat{C}'$, with our ground truth GT. We initialise trimaps using $D_u = s l_{\hat{C}} + 9$ pixels over various parameter values s and apply matting to produce an opacity mask where pixels in Ω_u are assigned an alpha value in the range $[0,1]$. We then threshold the mask at a value of 0.5 to achieve binarisation and extract the contour section associated with the fin boundary $\hat{C}'$ for all 120 training images.

We evaluate the resulting contours by computing Hausdorff and Chamfer distances to the ground truth. The results when using affinity matting can be seen in Figure 4, whilst Table 1 shows a benchmark and comparison with alternative methods. Overall, we find that using affinity-based matting at $s = 0.009$ leads to the best reconstructions in our domain, reproducing expert human labelling results with an average localisation error of 0.877 pixels.

Method	Accuracy (s) (pixels)	Robustness (s=0.016)(pixels)	Processing time (s=0.016)(seconds)
Affinity matte	0.877 (0.009)	1.001	64.43
Colour matte	1.970 (0.005)	>2.454*	>302.5*
GrabCut	1.366 (0.006)	1.9431	9.87

Table 1: **Accuracy Benchmarks.** Method comparison via statistics inspired by [16]: Accuracy is calculated as the smallest localisation error produced by a method for any value of s. Robustness is computed as the average localisation error at wide baselines of $s = 0.016$ and $*s = 0.013$.

5 Discussion

While the Hausdorff distance provides a useful measure of how localisation errors change with respect to s, it cannot be assumed that points most poorly localised initially by $\hat{C}$ corre-

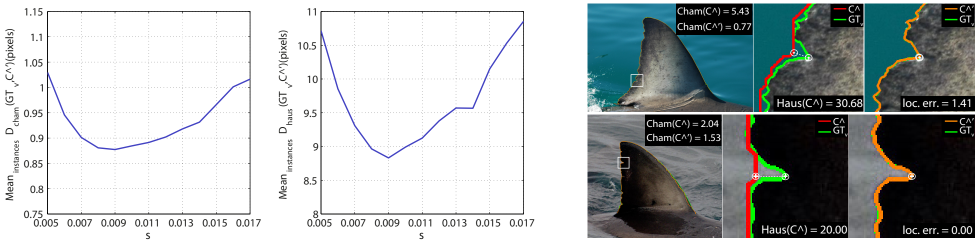

Figure 4: **Affinity Matting Performance. (Left)** Chamfer distances varying with respect to widths of the Ω_u, defined by the parameter s (abscissa) using the affinity matte approach. **(Middle)** Average maximum localisation errors. We see that by both measures, the optimal width of Ω_u corresponds to $s = 0.009$. **(Right)** Coarse initial contour $\hat{C}$ (red), ground truth GT (green) and $\hat{C}'$ (orange) after affinity based matting using $s = 0.016$. The central column depicts the ground truth locations for which localisation errors induced by $\hat{C}$ were greatest. Importantly, we see that not only are these extreme points well localised, but that the average localisation errors are also small.

spond to those most poorly localised by $\hat{C}'$. This is significant, since we observe that points least well localised by $\hat{C}$ correspond to prominent notches on the fin edge (Figure 4, right).

For the affinity-based matte, we observe that on average, these points are best localised when $s = 0.016$ (Figure 5, first), with an average localisation error of 2.3 pixels. Affinity matting is particularly sensitve to the trimap being initialised inaccurately and to points on GT being located outside of Ω_u (Figure 6, right). Meanwhile as s increases, the average distance between pixels close to the ground truth and the nearest labelled pixels will increase. This increases the likelihood of increased distance in colour space between labelled and unlabelled pixels, which in turn increases the likelihood of greater localisation errors. As measured by the Hausdorff distance, the optimal trade-off between these two effects occurs when $s = 0.009$ (see Figure 4, left). Affinity matting specifically outperforms colour sampling mattes in cases of low correlation between the labelled and unlabelled pixels (Figure 6, left) or overlap between the foreground and background colour distributions (Figure 6, middle). The affinity method encourages pixels with similar colours to take similar values, thereby exploiting the structure of the image within the unlabelled region.

In comparison to other methods tested, affinity matting also proved competitive regarding computational speed as shown in Figure 5 (rightmost). This can be explained by comparing the terms $\mathbf{X}_{Q_i}$ in Equation 3 and $\mathbf{X}_{N_i}$ in Equation 4. We use a 3×3 neighbourhood to compute $\mathbf{f}_{N_i}$ in the affinity based approach, meaning that in computing $\mathbf{f}_{N_i}$, we must solve for 9 unknowns. By contrast, in computing $\mathbf{f}_{Q_i}$ we solve for 160 unknowns, corresponding to the combination coefficients for 80 pixels in $Q_l^{f\prime}$ and 80 in $Q_l^{b\prime}$ respectively.

Readers may consider graph cut approaches [3] a viable alternative, we therefore also evaluate GrabCut [13] on the shark fin dataset. GrabCut minimises an energy functional E over target pixels in I:

$$E = U(\boldsymbol{\alpha}, I, \boldsymbol{\theta}) + V(\boldsymbol{\alpha}, I) \tag{7}$$

where $\boldsymbol{\theta}$ is a foreground-background model (GMMs with $k = 5$ used here) and $\boldsymbol{\alpha}$ is the binary assignment vector determining whether pixel i is foreground or background (see Rother et al. [13] for full details). GrabCut seeks a *globally* optimal segmentation based on a global colour model. Yet, since a coarse local initialisation is available in the case at hand, GrabCut's advantage by sampling globally is widely offset. We observe in the fin domain that GrabCut is sensitive to *global* initialisation changes (see Figure 7), whereas the mattes have a reliance on being intialised accurately *locally*. Moreover, we observe that the boundary

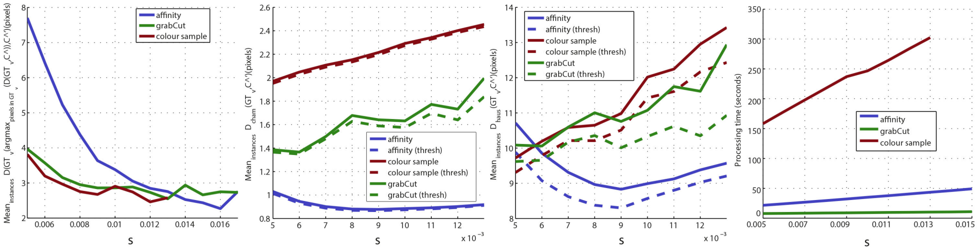

Figure 5: **Method Comparisons. (First)** Minimum distance from GT to $\hat{C}'$ for points on GT for which the minimum distance to a point on $\hat{C}$ was greatest (usually deep fin notches). Note, that affinity matting outperforms other methods for high s, yet under-performs for low s. **(Second)** Chamfer distance plotted against width of Ω_u. We see that the affinity method has the smallest localisation error for any value of s (is most accurate, see Table 1). **(Third)** The Hausdorff distance is also smallest for the affinity method, except when Ω_u is very narrow.

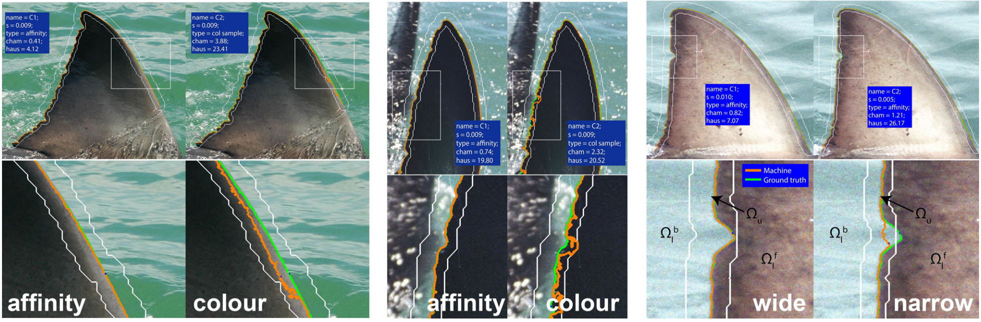
name = C1; s = 0.009; type = affinity; cham = 0.41; haus = 4.12
name = C2; s = 0.009; type = col sample; cham = 3.88; haus = 23.41
affinity
colour
name = C1; s = 0.009; type = affinity; cham = 0.74; haus = 19.80
name = C2; s = 0.009; type = col sample; cham = 2.32; haus = 20.52
affinity
colour
name = C1; s = 0.010; type = affinity; cham = 0.82; haus = 7.07
name = C2; s = 0.005; type = affinity; cham = 1.21; haus = 26.17
Machine
Ground truth
Ω_u
Ω_l^b
Ω_l^f
wide
narrow

Figure 6: **(Left/Middle)** Examples of affinity matting (left columns) producing reliable reconstructions whilst colour matting (right columns) fails. **(Right)** Example of narrow choice of D_u illustrates that affinity-based matting is robust, but not invariant, to changes in Ω_u.

term V in Equation 7, encourages pixels in regions of high contrast to take different values of alpha *independently* of $\boldsymbol{\theta}$. GrabCut thus has a tendency to segment the foreground fragmentedly or segment along background edges which exhibit higher contrast than that present at the true boundary location (see Figure 7). These two effects contribute to an overall decreased performance compared to affinity matting for the domain and dataset at hand (see Table 1).

6 Conclusion

In this paper we have investigated methods for reconstructing fin shape for great white sharks in a pixel-accurate manner given coarse pre-segmentations. In particular, we have evaluated the segmentation accuracy of affinity matting against colour matting and GrabCut using an expert-annotated dataset of great white shark fins, photographed in the wild.

For this domain we conclude that affinity mattes are able to obtain a more accurate overall classification result than direct colour sampling or GrabCut. The method was shown robust to differing initialisations including those encompassing the widest range of expected ground truth locations, enabling accurate segmentation of both fine as well as prominent notches on the fin outline. Overall, affinity mattes proved capable of localising ground truth contours to within-a-pixel. The method can be used as part of an ID system for great white sharks [7].

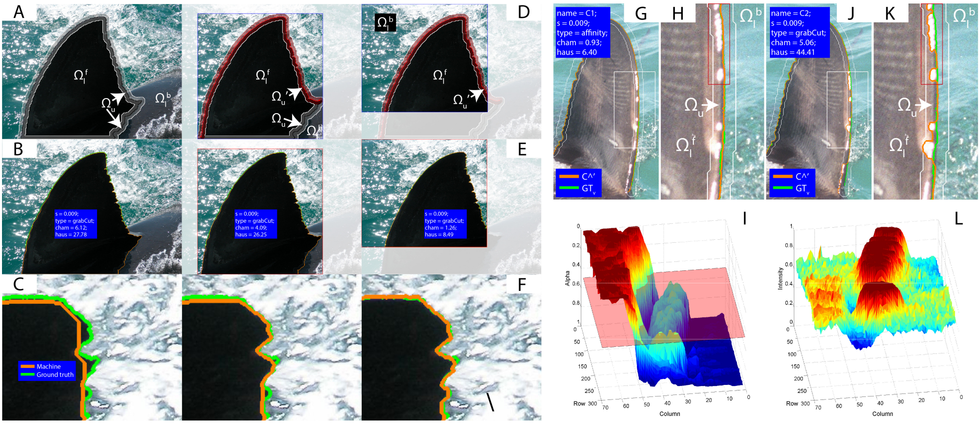

Figure 7: **GrabCut Sensitivity.** **(Left)** GrabCut initialised on the trimap **(A)** where the region labelled definite background (Ω_l^b) contains pixels representing the shark body. Fine details on the fin edge being missed **(B,C)**. By contrast, when initialised with a hand-drawn bounding box **(D)** errors can be avoided **(E,F)**. **(Right)** Illustration of segmentation differences between affinity-based matting **(G-I)** and GrabCut **(J-L)**.

Acknowledgements

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

References

[1] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(5):898–916, 2011.

[2] Pablo Arbeláez, Jordi Pont-Tuset, Jonathan T Barron, Ferran Marques, and Jitendra Malik. Multiscale combinatorial grouping. CVPR, 2014.

[3] Yuri Boykov and Gareth Funka-Lea. Graph cuts and efficient nd image segmentation. *International journal of computer vision*, 70(2):109–131, 2006.

[4] Chandan Gope, Nasser Kehtarnavaz, G Hillman, and Bernd Würsig. An affine invariant curve matching method for photo-identification of marine mammals. *Pattern Recognition*, 38(1):125–132, 2005.

[5] GR Hillman, B Wursig, GA Gailey, N Kehtarnavaz, A Drobyshevsky, BN Araabi, HD Tagare, and DW Weller. Computer-assisted photo-identification of individual marine vertebrates: a multi-species system. *Aquatic Mammals*, 29(1):117–123, 2003.

[6] Paul W Holland. Weighted ridge regression: combining ridge and robust regression methods, 1973.

[7] Benjamin Hughes and Tilo Burghardt. Automated identification of individual great white sharks from unrestricted fin imagery. In *BMVC 2015: Proceeding of the British Machine Vision Conference 2015*. BMVA Press, 2015.

[8] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.

[9] H Kühl and T Burghardt. Animal biometrics: Quantifying and detecting phenotypic appearance. *Trends in Ecology and Evolution*, 28(7):432–441, 2013.

[10] AD Marshall and SJ Pierce. The use and abuse of photographic identification in sharks and rays. *Journal of fish biology*, 80(5):1361–1379, 2012.

[11] Eric N Mortensen and William A Barrett. Interactive segmentation with intelligent scissors. *Graphical models and image processing*, 60(5):349–384, 1998.

[12] Elena Ranguelova, Mark Huiskes, and Eric J Pauwels. Towards computer-assisted photo-identification of humpback whales. In *Image Processing, 2004. ICIP'04. 2004 International Conference on*, volume 3, pages 1727–1730. IEEE, 2004.

[13] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (TOG)*, 23 (3):309–314, 2004.

[14] R Stanley. Darwin: Identifying dolphins from dorsal fin images. *Senior Thesis, Eckerd College*, 1995.

[15] AM Van Tienhoven, JE Den Hartog, RA Reijns, and VM Peddemors. A computer-aided program for pattern-matching of natural marks on the spotted raggedtooth shark carcharias taurus. *Journal of Applied Ecology*, 44(2):273–280, 2007.

[16] Jue Wang and Michael F Cohen. *Image and video matting: a survey*. Now Publishers Inc, 2008.

[17] Yuanjie Zheng and Chandra Kambhamettu. Learning based digital matting. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 889–896. IEEE, 2009.