Computer Assisted Photo Identification of *Dermochelys coriacea*

P.M. de Zeeuw, E.J. Pauwels,
Centrum Wiskunde & Informatica, Amsterdam (NL)
{*paul.de.zeeuw,eric.pauwels*}*@cwi.nl*

E.B. Ranguelova
Prime Vision, Delft (NL)
e.ranguelova@primevision.com

D.M. Buonantony
Duke University
Beaufort, North Carolina (USA)
danielle.buonantony@duke.edu

S.A. Eckert
WIDECAST
Ballwin, Missouri (USA)
seckert@widecast.org

1. Introduction

Identification marking is a fundamental tool used for conservation, management and research of wild animal populations. For marine turtles, marking is generally accomplished through the use of plastic or metal livestock tags attached to a flipper, or through the use of injected electronic Passive Integrated Transponder (PIT) tags. However, difficulties in applying long-term marking methods on sea turtles have significantly hindered the ability to accurately measure a wide variety of biological and population variables like population size, reproductive output and longevity. One method specifically evaluated on the leatherback sea turtle (*Dermochelys coriacea*) is the use of photo-identification to identify individual animals. Leatherbacks are listed as *critically endangered* by the International Union for Conservation of Nature and Natural Resources, see also Eckert et al. [4].

A feature unique to leatherbacks is the presence of the so-called pink spot, a de-pigmented patch located on the dorsal surface of the head, directly over the pineal gland, see Figure 1. Pioneering studies by McDonald and Dutton [6] showed that the shapes of pink spots appeared distinct enough between individuals to be used as a unique identifier.

However, as data collections expand, photo-identification by human analysts involves increasingly laborious and tedious browsing through a photo database, up to prohibitive levels. We therefore seek to construct a computer-assisted system, capable of automatically matching pink spot photographs against a database of earlier encounters, with the purpose of identifying individual and re-migrant leatherbacks. Because re-sightings are not frequent, one operational requirement is that the risk

of overlooking a genuine match should be extremely low. For related work, on other species, see Branzan Albu et al. [1], Ranguelova et al. [7], Van Tienhoven et al. [8] and references therein.

Figure 1. The same leatherback photographed at different occasions. Yellow rectangles indicate manual cropping prior to automatic matching, see Figure 2.

Overview Images of the same animal taken at different occasions may vary with respect to illumination,

resolution and viewing angle. Figure 1 highlights how manual cropping of the pink spot involves a somewhat arbitrary choice. Furthermore, through the years, the spot can become partially occluded due to scars and pollution. We have to design a robust image matching algorithm able to withstand the above variations. Since the head's dorsal surface is essentially flat and the size of the spot is small with respect to the distance to the camera, the geometric deformation between images of the same animal can be satisfactorily modeled by an affine transformation, i.e. a combination of translation, scaling, shearing and rotation.

The computer-assisted system that we propose is based on the Scale Invariant Feature Transform (SIFT, Lowe [5]) which selects so-called keypoints and gathers gradient information around them in descriptor vectors. To achieve the required low *false negative rate* we perform the subsequent matching of images while using relaxed control parameters. This results in many spurious matches which are then pruned by insisting on the affine consistency of the keypoint locations, see Figure 2 for an illustration. Additionally, we check the cross correlation of gray values at a number of selected high contrast regions. The components of the matching procedure are explained in Section 3. In Section 4 we put the procedure to the test by means of two datasets, one of which involves important changes over time.

2 Image Preprocessing

The first step when entering an image in the system is cropping, which is performed manually as part of the data entry procedure. This results in a smaller image, in which the pink spot features prominently and most of the background clutter has been removed, as illustrated by the yellow rectangles in Figure 1.

Although SIFT keypoint matching has been extended to colour images [2], it proves advantageous to convert our images into grayscale as they contain little colour information, carrying shades of pink and gray. We opt for the following two conversions:

1. Image Fusion   In general, image fusion seeks to combine salient information in two or more images of the same scene into a single highly informative image. Here we consider the R, G and B component as three separate images that are fused into one, the grayscale image. The particular (multiresolution) fusion method we use is described in De Zeeuw [3]. Hereby the conversion into grayscale appears to be without noticeable loss of information.

2. Opponent based colour contrast   To highlight the pink spot we compute the contrast (K) between the red component (R) and the (almost) average of green (G) and blue (B):

$$K = R - 0.4(G + B).$$

Notice that we settle for a factor 0.4 rather than 0.5 as it strikes a good balance between removing flashlight and enhancing contrast.

3 Image matching

SIFT keypoints extraction and matching   For a given grayvalue image the SIFT algorithm [5] searches — across different scales — for well-localized regions that stand out from their local background. It produces the location of the region centres, as well as the appropriate scale and the orientation with respect to the dominant gradients. Along with each keypoint p_i comes a *keypoint descriptor* δ_i, which is a feature vector (of length 128) summarizing local gradient information. These features are defined such that they are independent of image scaling and rotation and, to a considerable extent, invariant to changes in illumination and 3D camera viewpoint.

The similarity of images is then quantified by matching their SIFT keypoints, based on the corresponding descriptors. More formally, let's assume that the SIFT algorithm has identified n points p_i in I and n' points p'_j in I'. For each keypoint p_i $(i = 1 \ldots n)$ in I one looks for the best matching keypoint p'_j $(j = 1 \ldots n')$ in image I' by searching for the smallest distance $d(\delta_i, \delta'_j)$ between their respective 128-sized descriptors δ_i and δ'_j. Only if this minimal distance is significantly better than the second best match (to say, point p'_k with descriptor δ'_k) in the sense that $d(\delta_i, \delta'_j) < D_R d(\delta_i, \delta'_k)$ (for some threshold value $0 < D_R < 1$), then this match is retained; otherwise it is rejected.

The fraction D_R is called the *distance ratio*, and its value has a significant influence on the number of matches found. Low values for this control parameter make the inequality condition more exacting, resulting in a small number of keypoint matches, if any at all. Conversely, picking a D_R-value close to 1 relaxes the inequality condition, resulting in more point matches. The downside of the latter lenient threshold value is that the risk of producing spurious point matches between dissimilar regions in both images is also quite elevated, see top of Figure 2.

As mentioned before, due to the fact that resightings are relatively rare, the penalty for false negatives (overlooking an actual match) far outweighs

the inconvenience of a false positive (a match that does not hold up under scrutiny). For that reason we allow for lenient values of D_R (up to 0.9) and use additional checks on the geometry and grayvalues to remove erroneous matches. As a first sieve, we only retain matches that are *bi-directional*, i.e. they persist when we swap the roles of I and I'. This already significantly reduces the number of erroneous matches. Nevertheless, it is unlikely that all of them will be weeded out and we have to resort to the before-mentioned constraints (next).

Affinely consistent constellations At this point in the processing we have two paired-up sets, $P = \{p_1, \ldots, p_i, \ldots, p_k\}$ in image I and $P' = \{p'_1, \ldots, p'_i, \ldots, p'_k\}$ in image I', where each point p_i is uniquely matched to its counterpart p'_i. As the permissive choice of the D_R parameter frequently results in a multitude of spurious matches, we now search for the maximal subset(s) $S \subset P$, and the corresponding $S' \subset P'$, that are related through an affine transformation A,

$$A(p_i) = p'_i, \quad \forall p_i \in S.$$

The rationale is clear: since we know that the geometric deformation between different images of the same pink spot are well approximated by an affine transformation (see introduction), correctly matched keypoints need to satisfy the same constraint. Subsets S and S' will be called *affinely consistent constellations*. The larger the constellations, the stronger the evidence that the proposed matches are genuine.

To identify these constellations, we resort to exhaustive search. More precisely, we consider every possible quintet of keypoints in I together with the quintet of paired up keypoints in I' and check whether there exists an affine transformation that maps the quintet in I onto the quintet in I' (within a predefined tolerance). If this is the case, we try to expand the quintet by including more points for which the affine transformation proves accurate. This step is iterated, each time recomputing the affine transformation on the expanded set, until no more points can be added.

In this way, each initial quintet Q will give rise to affinely consistent constellations S_Q and S'_Q. The largest of such constellations (in terms of number of points) over all quintets Q are called the *largest affinely consistent constellations* of the paired-up pointsets P and P'.

Figure 2 illustrates how this procedure, when applied to the 31 matches generated by SIFT (top), removes erroneous matches to end up with affinely consistent constellations of seven points.
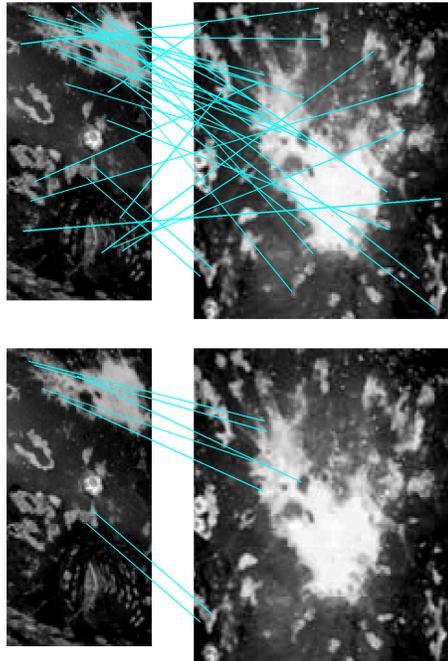
Figure 2. Matching the pineal spots of Figure 1. Top: Visualization of matched keypoints (31) using Lowe's matching algorithm. Bottom: Largest affinely consistent constellations: subsets of maximum size (7) that are related through an affine transformation.

Region-based checking At this point we are left with a constellation of (m say) affinely consistent matches. If this number is large (typically $m > 10$) then we can safely assume that the two images match. Conversely, if m is small ($m < 5$) then this a strong indication that the corresponding images are non-matching/different. If, however, m takes on an intermediate value then we need additional verification based on local image content.

We proceed as follows. Given the constellation of m affinely consistent bi-directional matches between the two images I and I' we can estimate the best fitting affine transformation A, i.e. the one that minimizes $\sum_{i=1}^{m} \|A(p_i) - p'_i\|^2$. We then apply this transformation to the whole image I, producing $A(I)$. If this transformation is for real, grayvalues in $A(I)$ and I' should be highly correlated. To check the above we select 15 high contrast subregions R' in I'. Using cross-correlation, we find the locations of the best matching counterparts (R) in the registered image $A(I)$. If the images I and I' are indeed matching

(and A is the correct affine transform) then the centres of the regions R and R' should be close. As the score for region-matching we count how many (r say) of the 15 selected regions satisfy the proximity. The final similarity score s for the image pair I and I' is then defined as the sum $s = m + r$.

4 Results

We test our system on two datasets. The first, which includes 76 images collected at Juno Beach, Florida (USA), is particularly challenging. Several individuals have been photographed at dates that are largely apart, both leading to pineal spots altered by pollution & scars and quite different photographing conditions. Moreover, after cropping several of the resulting pineal spot some images suffer from low resolution. We use this database to calibrate parameters in our overall matching algorithm. Setting the threshold for the similarity score s equal to 10 produces the best result. The separation between scores for matching and non-matching pairs looks well, though, unfortunately, one false negative remains. There are 4 false positives.

Dataset 1	$s \leq 10$	$s > 10$
non-matching image pairs	2822	4
matching image pairs	1	23

Next, we validate the procedure on a second image database numbering 151 images drawn from a large nesting population at Matura, Trinidad. Using the same threshold for the similarity score results in perfect classification.

Dataset 2	$s \leq 10$	$s > 10$
non-matching image pairs	11236	0
matching image pairs	0	89

5 Conclusions

The ability of the algorithms to define distinct points with the pineal spot for robust image matching demonstrates the effectiveness of our methodology and confirms the distinctiveness and reliability of the pineal spot as a unique identifier. These findings emphasize that photo-identification methodologies can be valuable alternatives to traditional tagging protocols. The expansion of photo-identification within the sea turtle field will allow researchers the ability to track population and biological parameters beyond nesting females, to include males and other age-classes of leatherbacks. The non-invasive and inexpensive nature of this technique also has the potential to increase the number of marking projects currently being undertaken. The cumulative benefits of this technique may have significant implications for leatherback conservation and management.

References

[1] A. Branzan Albu, G. Wiebe, P. Govindarajulu, C. Engelstoft, and K. Ovatska. Towards automatic model-based identification of individual sharp-tailed snakes from natural body markings. In *Proceedings of ICPR Workshop on Animal and Insect Behaviour (VAIB 2008), Tampa, FL, USA*, 2008.

[2] G. J. Burghouts and J. M. Geusebroek. Performance evaluation of local colour invariants. *Computer Vision and Image Understanding*, 113:48–62, 2009.

[3] P. M. de Zeeuw. A multigrid approach to image processing. In R. Kimmel, N. Sochen, and J. Weickert, editors, *Scale Space and PDE Methods in Computer Vision*, volume 3459 of *LNCS*, pages 396–407. Springer-Verlag, Berlin, 2005.

[4] K. L. Eckert and C. Luginbuhl. Death of a giant. *Marine Turtle Newsletter*, 43:2–3, 1988.

[5] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2(60):91–110, 2004.

[6] D. L. McDonald and P. H. Dutton. Use of pit tags and photoidentification to revise remigration estimates of leatherback turtles, dermochelys coriacae, nesting on st. croix, u.s. virgin islands, 1979–1995. *Chelonian Conservation and Biology*, 2(2):148–152, 1996.

[7] E. Ranguelova and E. Pauwels. Saliency detection and matching strategy for photo-identification of humpback whales. *ICGST International Journal on Graphics, Vision and Image Processing*, Special Issue on Features and Analysis, 2006.

[8] A. M. Van Tienhoven, J. E. Den Hartog, R. A. Reijns, and V. M. 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, April 2007.