# Leatherbacks Matching by Automated Image Recognition

Eric J. Pauwels[1], Paul M. de Zeeuw[1], and Danielle M. Bounantony[2]

[1] Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
{Eric.Pauwels,Paul.de.Zeeuw}@cwi.nl
[2] Nicholas School of the Environment and Earth Sciences, Duke University, USA
danielle.buonantony@duke.edu

**Abstract.** We describe a method that performs automated recognition of individual laetherback turtles within a large nesting population. With only minimal preprocessing required of the user, we prove able to produce unsupervised matching results. The matching is based on the Scale-Invariant Feature Transform by Lowe. A strict condition posed by biologists reads that matches should not be missed (no false negatives). A robust criterion is defined to meet this requirement. Results are reported for a considerable sample of leatherbacks.

## 1 Introduction

The ability to individually identify sea turtles in the field has been one of the most valuable tools in advancing our understanding of these animals. Marked or identified turtles allow for the measurement of a wide variety of biological and population variables (e.g. reproductive output, longevity, and survival rates). Traditional marking methods have included flipper, transponder, and mutilation tagging. In leatherbacks the pink spot, overlying the pineal gland on the dorsal surface of the head, has been reported as a unique identifier (McDonald and Dutton [6]). Leatherback nesting colonies of Trinidad offer the ideal research location for collecting photos of these spots as it annually supports nesting by 10,000 turtles. The Matura Beach/Fishing Pond nesting colony, located on the east coast of the island accounts for approximately half of all nesting on the island with over 150 turtles nesting per night. The beach is patrolled continuously by a local conservation organization, The Nature Seekers, which enabled most turtles to be detected. Photos are taken only during the laying stage of nesting to preclude disturbance of the turtle.

Identification of leatherbacks by humans involves laborious and tedious browsing through a (growing) photo database. Therefore, we seek to determine whether identification can be automated using image recognition algorithms. The time that can be put in watching colonies is limited, already for this reason the algorithm needs to avoid false negatives at all costs. The latter is an important issue for biologists, the presence of the same leatherback at a different place and different time should not be overlooked. The Scale Invariant Feature Transform

P. Perner (Ed.): ICDM 2008, LNAI 5077, pp. 417–425, 2008.

(SIFT, Lowe [4]) appears capable to provide us with automated matches robust to changes in 3D viewpoint and illumination, noise and occlusion.

Biologists are waking up to the possibilities of computer-assisted photo-identification and a number of stand-alone systems are under development (cf. [2,5,7,8]). The method demonstrated in this paper allows for a web based service by which one can query and contribute to a database of images. See [1] for a similar service under development also within the field of biodiversity.

The paper is organized as follows. In Section 2 we decribe the necessary preprocessing of images and the use of Lowe's SIFT features. Section 3 describes how to decide whether we can presume that images of pineal spots are matching or not. In Section 4 we provide statistics derived from a comparison with the groundtruth. There is the option of providing a future webservice, see Section 5.

## 2 Preprocessing and Feature Transform

### 2.1 Preprocessing

*Cropping.* An individual leatherback can uniquely be identified by its pineal spot [6], see the top row of Figure 1. We benefit from the fact that the pineal spot stands out in pink on the dorsal surface of the animal. We search and isolate the "pink spot" by human intervention, i.e. a rectangular region around the spot is selected. An additional advantage of the cropping is the reduction of dimensions which speeds up the subsequent processing. Clearly, the selection procedure introduces some arbitrariness as it is not always obvious to what extent "satellite" spots and marks should be included. However, as long as the main salient parts are retained, the resulting classification appears quite robust, see Section 4. The cropping is the only manual intervention required at this stage and typically takes 5 secs per image.

*Contrast enhancing.* The cropped colour image is turned into a gray-value image, where the gray-value is computed in such a way that it enhances the contrast between the pink spot and the dark background. This can be done adaptively (i.e. data-driven) by selecting the colour combination that corresponds to the first PCA (principal component analysis) factor. In the current implementation we simply convert a colour image into a gray-scale image by defining the gray value $K$ at each pixel as

$$K = R - 0.5(G + B)$$

where $R$, $G$ and $B$ are the intensity-values of the red (R), green (G) and blue (B) component.

### 2.2 Recapitulation on SIFT

To recognize (gray-value) images we use the features produced by the Scale Invariant Feature Transform (SIFT, Lowe [4]). This method selects so-called *keypoints* in an image. These are local points of interest, furnished with location, best fitting scale, and orientation with respect to the gradient. Along with

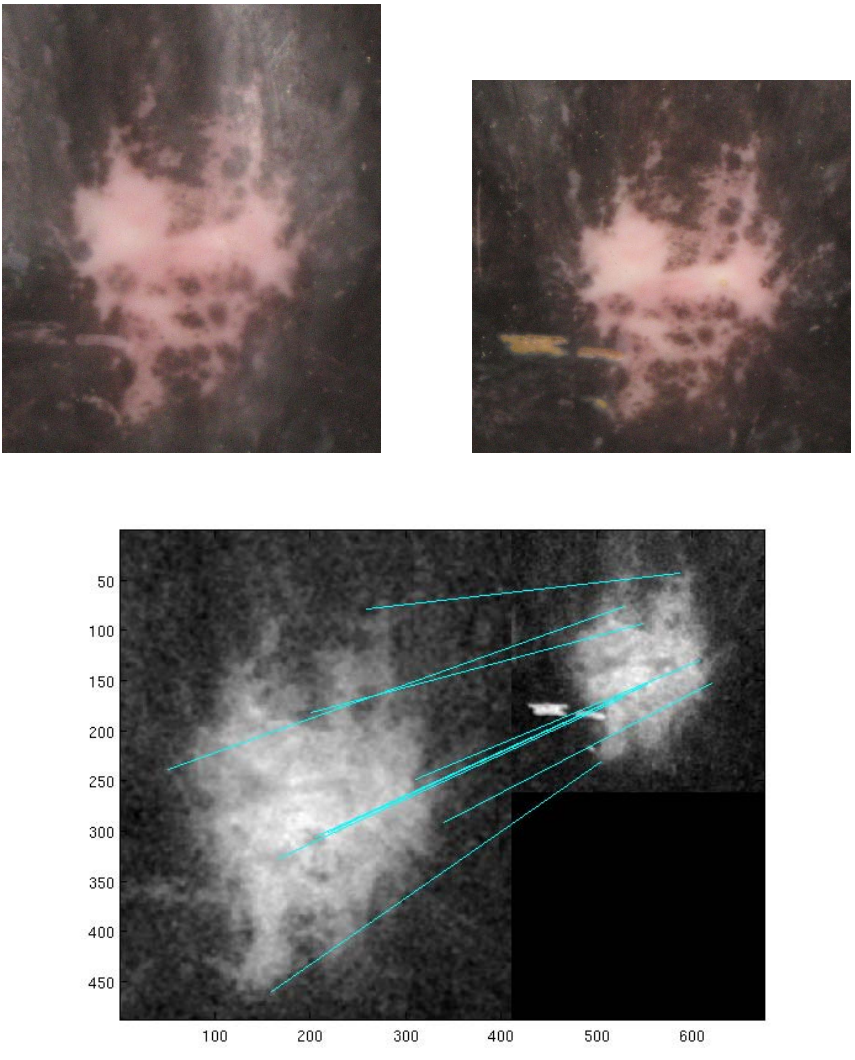


**Fig. 1.** Top: pineal spots of leatherbacks photographed on different days with different cameras. Bottom: matching keypoints found by SIFT.

each keypoint comes a *keypoint descriptor*, which is a feature vector summarizing local gradient information. The keypoints are selected in a strict manner through a cascade filtering approach. The features are defined such that they appear both invariant to image scaling plus rotation and, to a considerable extent, invariant to change in illumination and 3D camera viewpoint. Moreover, they are well localized in both the spatial and frequency domains, reducing the probability of disruption by occlusion, clutter, or noise. The descriptors prove highly distinctive, which allows a single feature to find its correct match with

good probability in a large database of features. Below we describe major stages within the intricate transform, and omit lots of (important) details.

1. **Extrema detection in scale-space.** The image $I(x, y)$ is convolved with a difference-of-Gaussian function which computes the difference of two nearby scales (separated by a constant factor $k$)

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \tag{1}$$

where

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}.$$

This can be computed efficiently and closely approximates the result as if a scale-normalized Laplacian of Gaussian $\sigma^2 \Delta G$ were applied (Lindeberg [3]).

In order to detect the local maxima and minima of image $D(x, y, \sigma)$, each sample point is compared not just to its eight direct neighbors but also to its nine direct neighbors in a scale above and below. It is selected only if it is larger than all of these neighbors or smaller than all of them. Still, the scale-space difference-of-Gaussian function has a large number of extrema, all candidate keypoints. Fortunately, a coarse sampling of scales suffices.
2. **Keypoint localization in scale-space.** A Taylor expansion (up to quadratic terms) of $D(x, y, \sigma)$ is used to determine an accurate location of the extremum in the coordinates $(x, y, \sigma)^T$. An expansion around the newly found extremum helps to detect low contrast, upon which the candidate keypoint is rejected as it is deemed unstable. Edges occurring in the original $I(x, y)$ provide another source of extrema with poorly defined locations. This situation is detected when the ratio of principal curvatures at an extremum rises above a certain threshold, upon which, again, the candidate keypoint is rejected as it is deemed unstable.
3. **Orientation assignment.** We proceed with the keypoints that have remained. In order to achieve rotation invariance for our keypoint descriptor to be (next stage, stage 4), we want to determine the keypoint orientation. The convolved version of $I(x, y)$ with scale closest to the one of the keypoint is selected for doing so. Magnitude and direction of the gradient are computed pixelwise using simple differences. A histogram is formed from the orientations of sample points within a certain Gaussian-weighted circular window around the keypoint. Obviously, peaks in the histograms correspond to dominant directions. At most two of such directions are taken into account (two directions leading to two different keypoints).
4. **Descriptor assignment.** This stage is similar to the previous one in that orientation histograms are computed. Again the scale of the keypoint determines the level of Gaussian blur for the image. To achieve rotational feature invariance, coordinates are rotated relative to the keypoint orientation as determined in the previous stage. The feature descriptor is computed as a set of orientation histograms over $4 \times 4$ sampling regions. Only 8 different orientations are considered, leading to 8 bins in each histogram. This leads to a feature vector / descriptor of $4 \times 4 \times 8 = 128$ elements per keypoint.

**Keypoint descriptor matching.** Comparing two images $I$ and $I'$ now boils down to comparing their respective sets of keypoints and descriptors. To decide whether an individual keypoint with descriptor in one image matches with a counterpart in the other image is not trivial. A uniform treshold on distance between descriptors is not wise as some descriptors discriminate more easily than others. For a positive match it is not good enough for mutual descriptors to be at close range. Far too many descriptors may apply, hereby invoking lots of false matches. Instead, a match of descriptors is required to *excel*. This is expressed by the criterion explained below. For keypoint $p_i$ in image $I$ one looks for the best matching keypoint $p'_j$ in image $I'$ by searching for the smallest distance $d(\delta_i, \delta'_j)$ between their 128-sized descriptors $\delta_i$ and $\delta'_j$. This point match will only be retained if it excels: the (minimum) distance of the first choice should be smaller than a predefined fraction of the second best choice. More formally, $p_i$ in image $I$ is matched to $p'_j$ in $I'$ only if

$$d(\delta_i, \delta'_j) = \min_k d(\delta_i, \delta'_k)$$

and

$$d(\delta_i, \delta'_j) < D_R \min_{k \neq j} d(\delta_i, \delta'_k).$$

Otherwise it is rejected which implies that keypoint $p_i$ has no match in $I'$. The fraction $D_R$ is called the *distance ratio* by Lowe [4] and is often fixed at a value of 0.6.

## 3 Matching of Images

Here we explain on what grounds (criteria) we presume the result of matching two images to be positive or negative and how reliable (and why) we want our presumptions to be. We rely on SIFT keypoints and use the accompanying descriptors. One needs to be aware that the matching of images is *not* symmetric: it depends on whether an image is considered a *query* or a *reference* image. The asymmetry is due to the way a match of descriptors has been defined (see the last paragraph of Section 2.2). A case in point is that one keypoint in the "query" image may resemble more than one keypoint in the "reference" image. To come up with a *symmetric* similarity measure we compute the number of bi-directional matches ($n_{bi}$): i.e. matches are only retained if they persist when swapping the roles of query and reference image. If a point-match is bi-directional the chances of it being erroneous are slim (see the lower part of Figure 1 for examples).

Deformations between different images that occur are due to the use of different cameras at different times by different people. This involves differences in resolution or scale, rotations and translations, changes in illumination (including glare), viewing angles and pollution (see the upper part of Figure 1 for examples). SIFT is apt to deal with such variations. However, since we cannot afford to overlook a genuine match, we relax the value of the distance ratio $D_R$

to 0.7 (see Section 2.2). The net result of this adjustment is that the number of matching keypoints between the query and reference image will be higher.

The standard way to decide whether two images are similar could be straightforward: compute the number of bi-directional matches $n_{bi}$ and compare it to a predefined threshold. The images are then declared to be either matching or non-matching depending on whether or not $n_{bi}$ exceeds this threshold. Again, as explained before, it is of paramount importance to reduce the risk of overlooking a genuine match. We therefore thread cautiously and introduce *two* thresholds: an upper threshold $n_{bi}^{high}$ and a lower one $n_{bi}^{low}$. If the number of bi-directional matches ($n_{bi}$) between two images exceeds $n_{bi}^{high}$ then we presume to have a high quality match between the images and it is kept in the database. If, on the other hand, $n_{bi} < n_{bi}^{low}$ then the images appear dissimilar and the match is rejected. For image pairs that achieve a score in between these two thresholds, this is substantial evidence that the images might be similar but it needs to be backed up by an additional test (introduced below).

The deformation between different images of the same spot is moderate (see above). We therefore assume that if keypoints in the query image are correctly matched to their counterparts in the reference image, the distance between any pair of keypoints in the query image should be the same (up to a scaling) as the distance between the corresponding points in the reference image. This can easily be checked by regressing the distances in the reference images over the corresponding distances in the query image. Data points due to correct point matches will trace out a line, the slope of which reflects the afore-mentioned scaling factor. Mismatches on the other hand, will create outliers.

The proposed additional test can now be summarized as follows: for two images, find all pairs of points $p_i$ (in the query image) and $p'_i$ (in the reference image) which are joined by a bi-directional match. Next, compute the distances between all such points in each image separately. This results in a set of distance $d_{ij} = d(p_i, p_j)$ for the points in the query image, and another set $d'_{ij} = d(p'_i, p'_j)$ for the corresponding points in the reference image. The latter set of values is regressed on the the former (using the regression model $y = ax + \epsilon$ which corresponds to a line that passes through the origin). As argued above, the fit of regression model reflects the quality of the match. This is quantified by computing the *mean squared error* (MSE) for the regression:

$$MSE = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

where $\hat{y}_i$ is the predicted value based on the regression. If the MSE exceeds a predefined threshold, the regression fit is low indicating the the point matches are erroneous. As a result the images are classified as non-matching. If on the other hand, the regression fit is satisfactory, we conclude that the point matches — although relatively few in numbers — enjoy a consistency that is indicative of true underlying similarity. The image pair is therefore tagged as a potential

match, to be verified by a human expert for final validation or rejection. The ones that are retained are presented to the user for a final confirmation or rejection decision.

## 4 Results

In a first experiment we worked with a database of 613 images that were collected over the period of about six weeks in the Leatherbacks nesting colonies of Trinidad. During the night, groups of around 150 turtles would emerge from the sea to lay and bury their eggs on the beach. During this activity the pineal spot of most animals was photographed twice, usually within the time span of a few minutes. As a consequence, the database comprises lots of individual animals for which we have two photos taken in quick succession and labeled to reflect the fact that they depict the same individual. These pairs are very valuable as they furnish us with a set of genuine matches that can be used to check minimal performance measures (e.g. whether the number of false negatives among these trivial matches is actually zero). In addition to these trivial matches, there are the more interesting repeat encounters where the same individual was photographed on different nights. In the current database 13 such individuals were discovered by manual inspection. The challenge faced by the matching algorithm outlined above therefore amounts to identifying all true matches (i.e. both the trivial and the non-trivial ones) while simultaneously minimizing the number of images that need to be checked manually.

Recall that the matching decision logic involves two thresholds ($n_{bi}^{high}$ and $n_{bi}^{low}$, see Section 3) for the number $n_{bi}$ of bi-directional matches. In the current experimental set-up we use the values $n_{bi}^{high} = 10$ and $n_{bi}^{low} = 3$. If the number of bi-directional matches between two images exceeds 10 then we presume a high level of similarity and they are automatically accepted as a matching pair. Conversely, if the number of matches is less than 3 then the image pair is automatically rejected. Finally, if $3 \leq n_{bi} \leq 10$ then we compute the square root of the MSE for the regression model. If $\sqrt{MSE}$ exceeds a threshold (which has been set equal to 7% of the data range), then the regression fit is deemed unsatisfactory and also this pair is rejected. If however $\sqrt{MSE}$ is smaller than this threshold value, the image pair is presented to a human supervisor for final approval or rejection.

The algorithm checked $613 \cdot 612/2 = 187,578$ image pairs. The above outlined decision strategy succeeded in recovering all true matches while no genuine matches were overlooked. Notably, the algorithm managed to uncover one additional match which happened to be overlooked by human experts. A total of 73 pairs (i.e. less than 0.04% of all pairs) were singled out by the algorithm for final inspection by a human supervisor. For the biologists involved this algorithm therefore provided highly reliable and welcome assistance.

**Overview algorithm and results.** Let $p_i \quad (i = 1, \ldots, n_{bi})$ be the keypoints in the query image ($Q$) that have been bi-directionally matched (using SIFT

descriptors) to keypoints $p'_i$ in the reference image ($R$), i.e. if $m_{AB}()$ denotes the matching function from image $A$ to image $B$, then $\forall i = 1, \ldots, n_{bi} : m_{QR}(p_i) = (p'_i)$ AND $m_{RQ}(p'_i) = (p_i)$. Hence, the number of bi-directional matches between images $Q$ and $R$ equals $n_{bi}$.

| `Algorithm` | | |
|---|---|---|
| `if` $n_{bi} > n_{bi}^{high}$ | | Accept match between images $Q$ and $R$; |
| `else if` $n_{bi} < n_{bi}^{low}$ | | Reject match between images $Q$ and $R$; |
| `else` | | Compute distances $d_{ij} = d(p_i, p_j)$ and $d'_{ij} = d(p'_i, p'_j)$, regress $d'_{ij}$ over $d_{ij}$ and compute $\sqrt{\text{MSE}}$; |
| | `if` $\sqrt{\text{MSE}} > q$ | Reject match between images $Q$ and $R$; |
| | `else` | Present presumed match between $Q$ and $R$ to human supervisor for final confirmation or rejection. |

In the current implementation $n_{bi}^{low} = 3$, $n_{bi}^{high} = 10$, and $q$ equals 7% of the $d'_{ij}$ range, i.e. $q = 0.07(\max\{d'_{ij}; j > i\} - \min\{d'_{ij}; j > i\})$. The results for the current database are summarized in the table below.

| | |
|---|---|
| Nr. of images | 613 |
| Nr. of false positives | 0 |
| Nr. of false negatives | 0 |
| Nr. of pairs processed | 187,578 |
| Nr. of pairs retained for manual inspection | 73 (i.e. 0.04%) |

## 5 Discussion and Future Directions

Leatherback turtles migrate over large distances and it would therefore be interesting to collect all data in a readily accessible global database. It seems to us that a web-based database running the proposed photo-identification algorithm could be an interesting addition to the current data repositories. Since the only manual work involved is the initial cropping of the pineal spot and, possibly, the acceptance or rejection of a small number of ambiguous matches, organizing this as a web-service would be rather straightforward. This way groups of biologists could easily share and compare data collected at different times and locations. At the same time, it would allow large groups of amateurs to significantly contribute to the scientific enterprise by submitting their own pictures. We believe that this type of web-enabled collective effort will play an increasingly important role in the near future.

## Acknowledgments

We thank Scott A. Eckert (WIDECAST, Duke University) for sharing his expert knowledge on leatherback turtles. Danielle M. Bounantony acknowledges the practical support provided by The Nature Seekers, and the financial support received from the Environmental Internship Fund, Andrew W. Mellon Foundation and the Duke Center for Latin American and Caribbean Studies. This work was partially supported by MUSCLE, part of EU's Sixth Framework Programme for Research and Technological Development (FP6).

## References

1. de Zeeuw, P.M., Ranguelova, E., Pauwels, E.J.: Towards an Online Image-Based Tree Taxonomy. In: Perner, P. (ed.) ICDM 2007. LNCS (LNAI), vol. 4597, pp. 296–306. Springer, Heidelberg (2007)
2. Hillman, G., et al.: Computer-assisted photo-identification of flukes using blotch and scar patters. In: Proceedings of 15th Biennial Conference on the Biology of Marine Mammals (December 2003)
3. Lindeberg, T.: Scale-space theory: A basic tool for analysing structures at different scales. Journal of Applied Statistics 21(2), 224–270 (1994)
4. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
5. Mizroch, S., Beard, J., Lynde, M.: Computer Assisted Photo-Identification of Humpback Whales. In: Hammond, P., Mizroch, S., Donovan, G. (eds.) Individual Recognition of Cetaceans, pp. 63–70. International Whaling Commission, Cambridge (1990)
6. McDonald, D.L., Dutton, P.H.: 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. Ranguelova, E., Pauwels, E.J.: Saliency Detection and Matching Strategy for Photo-Identification of Humpback Whales. In: International Conference on Graphics, Vision and Image Processing - GVIP 2005, Cairo, Egypt, December 2005, pp. 81–88 (2005)
8. Van Tienhoven, A., den Hartog, J., Reijns, R., Peddemors, V.: 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)