www.icgst.com

Saliency Detection and Matching for Photo-Identification of Humpback Whales

E. Ranguelova, E. Pauwels
Center for Mathematics and Computer Science (CWI),
Kruislaan 413, 1098 SJ Amsterdam, The Netherlands,
[Elena.Ranguelova, Eric.Pauwels]@cwi.nl,
http://www.cwi.nl/pna4

Abstract

This paper presents a system for computer-assisted identification of humpback whales. Individuals of this species can be uniquely recognised by the light and dark pigmentation patches on their flukes (tails) and various acquired scratches. Identification is of great importance to biologists who need to match each new image to large collection of existing photographs. We have developed computer system which assists the user to perform the photo-matching faster. This is achieved by first segmenting the animal's tail from the sea and fitting an affine-invariant grid to it, followed by extraction and comparison of the segmented fluke patterns. The focus of this work is on a new binary salient pattern detector based on morphological operators. Each pattern is represented via five points in respect to the grid. We present an algorithm for matching the groups of interest points. The saliency detection and matching are general-purpose tools and can be used for other marine mammals and in other applications. The retrieval system using the developed strategy is tested on a humpback whale database with ground truth provided by an expert and demonstrates excellent performance.

Keywords: *Saliency detection, Pattern matching, Affine invariance, Photo-identification, Image retrieval, Humpback whales.*

1 Introduction

Recognition of individual marine mammals, (i.e. whales, dolphins and porpoises) is of great interest to marine biologists. It plays an important role in their long-term studies of the population and behavioural patterns of the mammals [1, 8]. Due to practical and legal issues associated with tagging the animals, the method of photo-identification offers a valuable non-invasive alternative. This approach utilises the unique natural markings which can be captured by photographing the dorsal fins or flukes (i.e.tails). Marine biologists discovered more than 30 years ago that the mammals exhibit sufficient variation in their natural markings to allow the identification of individuals. The markings in humpback whales are the white/light grey patches as well as some salient patterns: markings and long-lasting scars on the undersides of their flukes as illustrated on figure 1.

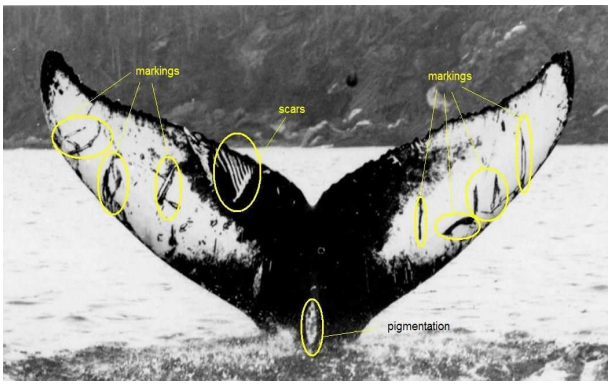

Figure 1: Humpback whale flukes. Some salient patterns that allow identification are highlighted.

As the photographic collections grew, so did the need for more efficient retrieval methods that would allow a researcher to quickly compare and match new photographs against a database of registered individuals. There are several approaches for photo- identification found in the literature. In [8] manually generated code is used, based on a set of 38 generic fluke patterns and which takes into account the shape of the central notch and the location of patterns. After the coding the resulting descriptors are subsequently used for matching. A curve matching technique, originally developed for the identification of bottle-nose dolphins [3] has been extended for the encoding of the fluke's

trailing edge of humpback whales in [1]. Recently, another affine-invariant curve matching method has been proposed in [5] for photo-identification of dolphins, sea lions and grey whales. In [7] affine moment invariants computed for the patches extracted from fluke images are used for identification of grey and humpback whales. Affine moment invariants from blotch and scar patterns have also been used in [6].

Previously, we proposed constructing an affine-invariant grid that is automatically fitted to the fluke, dividing it into relatively small invariant regions. Next, each region is characterised by the relative contribution of dark and light patches. This maps the visual information into a numerical feature vector which can then be compared to the feature vectors obtained from other images [9].

In this paper we introduce retrieval method based on comparison of the salient patterns. There are several reasons for choosing approach utilising patterns. They are the most natural visual characteristics used by the experts and it is a challenge to model the human visual recognition. Also, most of the current research efforts have been directed towards matching the fin or fluke contour, being the other important identification feature. Finally, the problem of analysing the salient content for the purposes of image database (DB) retrieval is a more generic one, and can be found in other applications. Our research goal in the photo-identification context is to develop saliency analysis tools which will reduce significantly the time used for manual search in large DB.

The remainder of the paper is organised as follows. Section 2 briefly outlines our previously proposed method for affine-invariant grid region features. The pattern matching approach is described in detail in section 3. The idea is to automatically extract salient patterns and to compare their affine-invariant representation by point matching. The retrieval setup is given in section 4 while the results and conclusions are presented in sections 5 and 6.

2 Retrieval Based on Simple Features

Previously, we have proposed a method for retrieval comprised of three main stages: image segmentation, affine-invariant grid fitting and matching based on simple grid region features [9].

2.1 Tail and Patch Segmentation

In the first stage the flukes are extracted form the background (sea, sky, etc.). Because the photographic quality is quite challenging, a semi-automatic segmentation has been adopted for extraction of the tail contour. The user initializes the process by specifying a rough initial contour (marker) within the tail. Marker-controlled watershed transformation is then applied to the modified image gradient automatically producing an estimated boundary contour for the tail. If needed, the interface allows the user to fine-tune the result. The program also allows morphological filtering for noise reduction. An optimal grey-level thresholding applied within the extracted contour obtains the patches. At the end of this stage, the tail with the patches and the identifying patterns is represented as binary image.

2.2 Affine-Invariant Grid

The images of marine mammal flukes often exhibit large variation in viewing angles, distance from camera and fluke inclination. It can be argued that since the tail surface is nearly planar with dimensions significantly smaller than the distance to the camera, these variations can be modelled using affine transformation (rotation, translation, scaling and shear) [7]. To make the matching robust to this variability, we have proposed an *affine-invariant coordinate grid*. Similar ideas for construction of invariant patch for a spotted dolphin identification have been employed in [4].

The grid is defined by the relatively stable anatomical landmarks- the left and right fluke tips (L and R on Figure 2) and the middle notch (O), currently indicated to the system by the user. These landmarks have been used also in [6, 5].

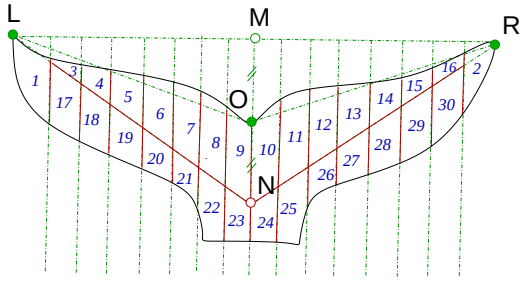

Figure 2: Affine grid construction and region labelling.

The base of the triangle $\triangle LOR$ is divided in two equal parts by the point M. The symmetrical point of M in respect to O, i.e. N is found. Each fluke is then divided into n parts with lines parallel to the median NM. Thus, the grid delineates $N_R = 4n - 2$ (the tips are considered single regions) grid regions. These regions are labelled 1 through N_R starting from the fluke tips and then scanning left to right, top to bottom. In the current implementation $n = 8, N_R = 30$ (Fig. 2).

Since the construction is based on affine invariant concepts (i.e. middle point, symmetry, equal distances, parallel lines), the resulting grid is invariant under affine transformations.

Figure 3 illustrates the segmentation and grid fitting to a pair of images of the same individual. Note that the salient patterns appear in the same grid position (for example in region 1, between regions 1 and 17 and in region 2) invariantly of the affine distortions.

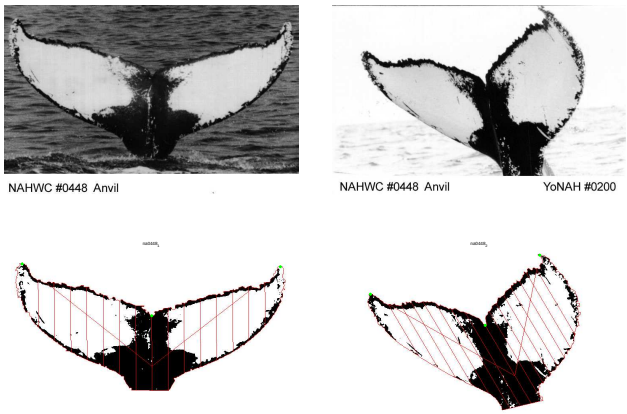

Figure 3: Fluke and patches segmentation of a pair of images of the same individual. Affine invariant grid fitting.

2.3 Feature Comparison

After the grid fitting, a simple N_R- dimensional feature vector $\mathbf{f} = (f_1, \ldots, f_{N_R})$ is computed. Each element f_i equals the ratio of the number of white pixels to the total number of pixels in the i-th grid region R_i. The feature vector for each image is stored in a database of features $\mathbf{F} = \{\mathbf{f}_1, \ldots, \mathbf{f}_N\}$ for all N images of the image DB. The matching involves comparison of the feature vector $\mathbf{q}$ for a query image against all entries in $\mathbf{F}$. The similarity is quantitatively measured as the average Euclidean distance per fluke segment:

$$d_{BW}(\mathbf{q}, \mathbf{f}) = \frac{\sqrt{\sum_{i=1}^{N_R} I_i (q_i - f_i)^2}}{\sum_{i=1}^{N_R} I_i}, \qquad (1)$$

where the indicator variable I_i indicates if the corresponding region of any of the pair of tails to compare should be considered, i.e.: $I_i = I_i^{\mathbf{q}} I_i^{\mathbf{f}}$. The indicator for each fluke and its corresponding grid equals 1 for all non-occluded regions and 0 for the ones which are occluded.

Marine biologists also use the fluke type when performing manual identification. The type T is defined as the percentage of white pixels in relation to the fluke area. There are 5 types of flukes, for example there are between 80% and 100% white pixels for a type $T = 1$, while there are between 0% and 20% - for a type $T = 5$. During comparison only relevant tails of types T_q, $T_q - 1$ and $T_q + 1$ (when applicable) are matched to a query tail of type T_q.

This retrieval method would be referred to as retrieval based on black and white (BW) ratio features.

3 Retrieval Based on Salient Patterns

A human observer can recognise an individual animal by a few visually salient patterns, while it is a difficult task for a computer vision system. It can be simplified by considering saliency in the case of binary images.

3.1 Binary Salient Pattern Detection

Consider binary image $\mathbf{B} \subset \mathcal{Z}^n \rightarrow \{0, 1\}$ and let $\mathbf{x} = 1$ correspond to a white pixel. Let $\mathbf{B}_1$ be the set of all white pixels in $\mathbf{B}$, i.e. $\mathbf{B}_1 = \{\mathbf{x}|\mathbf{B}(\mathbf{x}) = 1\}$ and $\mathbf{B}_0$ -the set of all black pixels $\mathbf{B}_0 = \{\mathbf{x}|\mathbf{B}(\mathbf{x}) = 0\}$. The notation $\widetilde{\mathbf{B}^1}$ stand for $\mathbf{B}^1$ without 'holes'. Let $\mathcal{B}^1 \subset \mathbf{B}^1$ be a large white connected component with area (number of pixels) proportional to the area of the image by a factor of Λ: $\text{card}(\mathcal{B}^1) \geq \text{card}(\mathbf{B})/\Lambda$.

The task of a salient detector is to find all salient patterns $\mathbf{S} = \mathbf{S}^i \cup \mathbf{S}^b$, which we define to be of two possible types. Firstly, there are the *inner* patterns $\mathbf{S}^i = S_{01}^i \cup S_{10}^i$, where S_{01}^i denotes all black regions entirely on white background and S_{10}^i- vice versa. We would like to make a distinction between small regions, which are considered as inner patterns themselves and large regions. On the borders of the latter we define the other salient pattern type as the major *border* irregularities between $\mathcal{B}^1$s and their black background (or vise versa) $\mathbf{S}^b = S_{01}^b \cup S_{10}^b$. The pixels in S_{01}^b have a value of 0 and the background is 1 and vice versa for S_{10}^b.

We propose a solution using *mathematical morphology* (MM) [10]. Let E be a morphological structuring element (SE). and $\check{E}$ is the reflected SE, i.e. $\check{E} = \{-e|e \in E\}$. If E is chosen to be symmetrical, for instance a disk with radius r (which can be selected depending on the dimensions of $\mathbf{B}$), then $\check{E} = E$ in respect to the centre of the disk. The two main MM set operations are *erosion* and *dilation*. The erosion of $\mathbf{B}$ with E is defined as $\epsilon_E(\mathbf{B}) = \min_{e \in E} \mathbf{B}(\mathbf{x} + e)$ and dilation as $\delta_E(\mathbf{B}) = \max_{e \in E} \mathbf{B}(\mathbf{x} + e)$. Intuitively, dilation expands an image object and erosion shrinks it. Using erosion and dilation two other important MM transformations can be constructed: *Opening* is defined as erosion followed by dilation $\gamma_E(\mathbf{B}) = \delta_{\check{E}}[\epsilon_E(\mathbf{B})]$ and *closing* is dilation followed by erosion $\phi_E(\mathbf{B}) = \epsilon_{\check{E}}[\delta_E(\mathbf{B})]$. Opening generally smooths an object contour, breaking narrow isthmuses and eliminating thin protrusions. Closing tends to narrow smooth sections of contours, fusing narrow breaks and long thin gulfs. The *white top hat (WTH)* operation gives the difference between the original image and its opening:

$$WTH_E(\mathbf{B}) = \mathbf{B} - \gamma_E(\mathbf{B}). \qquad (2)$$

It extracts all structures that cannot contain the SE and can be used to detect salient patterns of type S_{10}^b. On the other hand, the *black top hat (BTH)* gives the

difference between the closing of an image and the image itself:

$$BTH_E(\mathbf{B}) = \phi_E(\mathbf{B}) - \mathbf{B} \tag{3}$$

and can be used as detector for S^b_{01}.

We define a *saliency operator* $\mathbf{S} = \rho(\mathbf{B})$ by:

$$\mathbf{S}^i = \rho^i(\mathbf{B}) = \rho^i_{01}(\mathbf{B}) \cup \rho^i_{10}(\mathbf{B}) \tag{4}$$

$$\mathbf{S}^b = \rho^b(\mathbf{B}) = \rho^b_{01}(\mathbf{B}) \cup \rho^b_{10}(\mathbf{B}), \tag{5}$$

the inner patterns and the border irregularities operators respectively. They can be further expanded as:

$$S^i_{01} = \rho^i_{01}(\mathbf{B}) = \widetilde{\mathbf{B}^1} \cap \mathbf{B}^0 \tag{6}$$

$$S^i_{10} = \rho^i_{10}(\mathbf{B}) = \mathbf{B}^1 \setminus \bigcup_j \mathcal{B}^1_j \tag{7}$$

$$S^b_{01} = \rho^b_{01}(\mathbf{B}) = \bigcup_j BTH_E(\mathcal{B}^1_j) \tag{8}$$

$$S^b_{10} = \rho^b_{10}(\mathbf{B}) = \bigcup_j WTH_E(\mathcal{B}^1_j). \tag{9}$$

According to (6) the black regions on white background are obtained by intersecting the set of all black pixels with the set of all white without 'holes', i.e. the black 'holes' are detected. Equation (7) states that the small white regions on black background contain all white pixels which are not in any large white component $\mathcal{B}^1_j$. Equations (8) and (9) show how the border patterns can be extracted using the two top-hat transforms as defined by (2) and (3). For removing noise or tiny irrelevant objects the *area opening* operator γ_λ can be used to filter out all connected components whose area is smaller than λ [10].

$$\gamma_\lambda = \bigvee_i \{\gamma_{C_i} | C_i \text{ is connected and } card(C_i) = \lambda\},. \tag{10}$$

where $\bigvee$ is a point-wise maximum operation. Therefore, the expression for the salient operator becomes:

$$\rho = \gamma_\lambda \circ (\rho^i_{01} \cup \rho^i_{10} \cup \rho^b_{01} \cup \rho^b_{10}), \tag{11}$$

where the detectors are defined by equations (6) through (10).

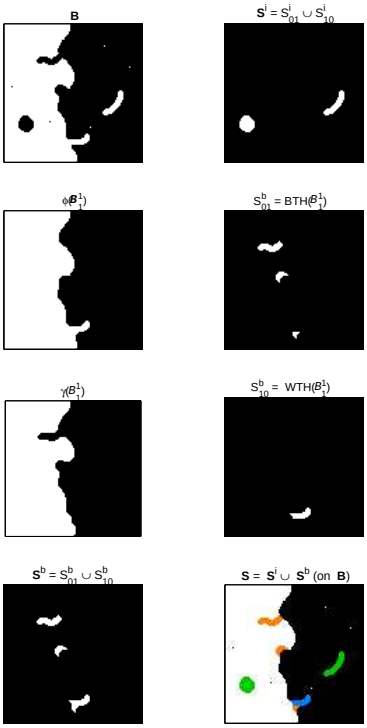

Figure 4: Binary salient patterns detection.

Figure 4 illustrates the saliency operator on a synthetic binary image with dimensions 100×100 pixels. The radius of the disk SE is $r = 5$ pixels. The other parameters are: $\Lambda = 100, \lambda = 3r = 15$. The sub-figures show the result of the different morphological operators. In the last sub-figure, the detected patterns are shown overlapping the original image and are colour codded: the inner patters -in green, the patterns of type S^b_{01} -in orange and S^b_{10} -in blue. Note also, that any small isolated regions are considered irrelevant as result of the area opening.

3.2 Pattern Representation

After each salient pattern has been extracted it is represented by a set of five characteristic points of the equivalent ellipse (up to the normalised second order moments). These points are the centroid C and the four points at the end of the major AB and the minor DF axes as shown in Figure 5. All coordinates of the points are converted from the image coordinate system to the affine coordinate system defined by L, O and R (Figs. 2 and 5). In the affine coordinate system the landmarks have coordinates $(u_O, v_O) = (0, 0)$, $(u_L, v_L) = (0, 1)$, $(u_R, v_R) = (1, 0)$ and the coordinates of the point C for example are defined as $u_C = |OC'|/|OL|$ and $v_C = |OC''|/|OR|$, where the length of any line segment is in respect to the image coordinate system $|PQ| = \sqrt{(x_P - x_Q)^2 + (y_P - y_Q)^2}$. The points C' and C'' are obtained by constructing CC' and CC'' parallel to OR and OL respectively.

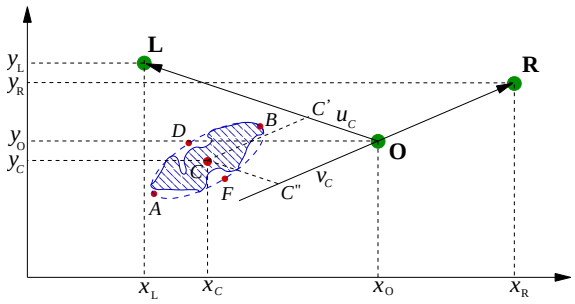

Figure 5: Pattern representation in affine coordinate system.

For each salient pattern S the grid region number R where its centroid C belongs to, is also stored. Therefore each pattern is represented as the following vector: $S = (R, u_C, v_C, u_A, v_A, u_B, v_B, u_D, v_D, u_F, v_F)$.

3.3 Pattern Matching

In order to facilitate matching between the extracted patterns we need a strategy for finding correspondence between the representing point-sets. The task is similar to the one of finding correspondence of interest points [2] in stereo-vision or in optical flow problems. Similarly, correspondences between a fixed number of points sampled from a contour are used for shape retrieval in [11]. Here, we propose an algorithm which works on the patterns' centroids and the criteria for similarity are the geometrical properties of the four region extrema (A, B, D, F (Fig. 5)).

Let $\mathcal{SQ} = \{S_i\}, i = 1, \ldots, m$ is the set of salient patterns extracted from the query image and $\mathcal{SD} = \{S_j\}, j = 1, \ldots, n$ (generally $m \neq n$) is the set of salient patterns from the DB image to be matched. The notation $< S_i, S_j >$ stands for a potential correspondence between the patterns S_i and S_j and the correspondence vector $\mathbf{d}_{ij}$ is the one that maps the patterns' centroids $C_i = (u_i, v_i)$ to $C_j = (u_j, v_j)$, i.e. $C_j = C_i + \mathbf{d}_{i,j}$, $\mathbf{d}_{i,j} = (u_j - u_i, v_j - v_i)$. The probability of $< S_i, S_j >$ is p_{ij} and some patters might not have a match with probability p^*.

The task is to find the best correspondence between the pattern sets $\mathcal{SQ}$ and $\mathcal{SD}$. We propose an iterative matching strategy which tries to find the best global correspondence based on local neighbourhood consistency.

Two patterns are considered *potentially corresponding* if their centroids belong to the same or neighbouring grid regions:

$$R_i = R_j \text{ or } R_j \in \text{neighbours}(R_i) \tag{12}$$

and if the distance between their centroids is below certain threshold:

$$|C_i - C_j| \leq d_{max}. \tag{13}$$

The *dissimilarity* between the patterns is defined as a weighed sum of the distances between the respective points:

$$\begin{aligned} ds_{i,j} = |C_i - C_j| + w_{AB}(|A_iB_i| - |A_jB_j|) \\ + w_{DF}(|D_iF_i| - |D_jF_j|), \end{aligned} \tag{14}$$

where w_{AB} is chosen larger than w_{DF} to facilitate higher importance of the major axes.

Each correspondence has a *weight* such that, the higher the weight the better the correspondence:

$$w_{i,j} = \frac{1}{1 + ds_{i,j}}. \tag{15}$$

Some patterns in one of the sets might have no match in the other. For any pattern, the correspondence with the highest weight should be the correct one, hence the probability that S_i has no correspondence with any S_j is

$$p_{i,j}^{*(0)} = 1 - \max_j w_{i,j}, \tag{16}$$

for all valid correspondences at the initialisation step. Two correspondences $< S_i, S_j >$ and $< S_k, S_l >$ within a neighbourhood specified by V are said to be *consistent* if

$$||\mathbf{d}_{i,j} - \mathbf{d}_{k,l}|| \leq d_{dif}. \tag{17}$$

The algorithm iteratively updates the probabilities of each potential correspondence depending on the probabilities of the neighbouring correspondences:

1. Detect all potential correspondences $< S_i, S_j >$, for all patterns in the query image $\forall S_i \in \mathcal{SQ}$ with all patterns in the database image $\forall S_j \in \mathcal{SD}$ subject to (12) and (13).

2. For each pattern in the query image $\forall S_i \in \mathcal{SQ}$ compute the vectors $\mathbf{d}_{i,j}$ for all potential correspondences detected at step 1. and the initial probabilities:

$$p_{i,j}^{(0)} = \frac{w_{i,j}}{\sum_{l \neq j} w_{i,l}}(1 - p_{i,j}^{*(0)}),$$

using (14), (15) and (16).

3. Iterate N_{it} times updating the probabilities:

$$\hat{p}_{i,j}^{(it)} = p_{i,j}^{(it-1)}(k_1 + k_2 q_{i,j}^{(it-1)}), \hat{p}_{i,j}^{*(it)} = p_{i,j}^{*(it-1)}, \tag{18}$$

where k_1 and k_2 are constants regulating the convergence speed. The quality of the correspondence is computed by:

$$q_{i,j}^{(it-1)} = \sum_{i \neq k} \sum_{|C_i - C_k| \leq V} p_{k,l}^{(it-1)}$$

for all correspondences $< S_k, S_l >$ consistent with $< S_i, S_j >$ (17). Normalise:

$$p_{i,j}^{(it)} = \frac{\hat{p}_{i,j}^{(it)}}{\sum_{\forall l} \hat{p}_{i,l}^{(it)}}$$

3.4 Similarity Score

After the correspondences between the salient patterns have been established, we propose the following *salient pattern similarity* score between two images:

$$s_{SP} = (\Sigma_w + \Sigma_{w_l})(s_M + s_I + s_{IL}), \qquad (19)$$

where Σ_w is the sum of all weights for correspondences with nonzero probabilities, Σ_{w_l} is the sum of all such weights greater than w_{thresh}. These terms reflect the quality of the found correspondences between the patterns. The other terms in (19) estimate quantitatively the similarity between the numbers of matched patterns and the actual number of patterns. They are the mutual similarity s_M, the similarity per image s_I and the term for similarities between matches with large weights per image s_{IL}:

$$s_M = \frac{1}{|m-n|+1}$$
$$s_I = \frac{1}{|m-N_T|+1} + \frac{1}{|n-N_T|+1}$$
$$s_{IL} = \frac{1}{|m-N_{T_l}|+1} + \frac{1}{|n-N_{T_l}|+1},$$

where N_T is the total number of detected correspondences between the two images and N_{T_l} is number of correspondences with weights higher than w_{thresh}.

4 Combined Retrieval

While the very fast BW method (section 2) uses simple features, the method based on matching salient patterns (SP), presented in section 3 is more robust, but more computationally intensive. To benefit from the advantages of both methods we propose to combine them. The combined (CB) method utilises SP matching only among the top relevant images found initially by the BW method.

The potential matches, which are finally presented for visual inspection by a human expert, are ranked using the following similarity:

$$s_{CB} = \frac{s_{SP}+1}{d_{BW}}, \qquad (20)$$

which is proportional to the similarity score of the SP method s_{SP} (19) and inversely proportional to d_{BW} (1). The nominator in (20) ensures that if the query image has no salient patterns, the BW ranking will be used as a final retrieval result.

5 Results

The proposed methodology was tested on a DB of $N = 340$ images of humpback whale flukes corresponding to 150 animals and each individual has from 2 up to 4 different images. The ground truth for the DB has been provided by an expert.

It is important to segregate the data based on their *photographic* (focus, angle, distance) and on their *recognition* (distinctive patterns) quality [8]. The quality codes used are: 1- excellent, 2- good to moderate and 3-poor. There are 292 images with photographic quality 1 or 2, 315- with recognition quality 1 or 2 and 275 with both qualities determined as excellent or good. Images of poor photographic quality are 48 (14% of the DB), while 25 (7%) have poor recognition quality and 8 (2.4%) images were coded as having both qualities of level 3.

All methods were implemented in MATLAB and executed on a Pentium, 2.4 GHz, 512 MB computer. The semi-automatic segmentation of the original images and the grid fitting [9] took a few minutes per image. The computation of the whole BW feature database **F** took about 40 minutes. The salient pattern detection took few seconds per image. This is a pre-processing stage for the existing photographic DB and can be done offline. When a new image is submitted to the system, after the pre-processing, it can be submitted as a query to the retrieval system.

5.1 Salient Pattern Detection

Figure 6 illustrates the performance of the salient patterns detector.

The same colour coding as for Fig. 4 is used. It can be seen from the middle sub-figure that the detector has captured all potential salient patterns from the segmented image. For this image of dimensions 1050×750, the elapsed time for the detection was 5.5 s. Due to the large number of potential salient patterns usually detected in a natural image, the system provides an option for the user to select a subset of the most significant patterns as illustrated.

On Figure 7 screen-shots from the system at the stage of the automatic salient pattern detection on another image are presented. Again the detector picks up all salient pattern structures as available after the binary segmentation.

5.2 Pattern Matching

After the patterns for the query and all DB images were available, the algorithm from section 3.3 was applied to their affine point-set representations. The speed of the pattern matching depends on the choice of parameters such as d_{max}, N_{it} and the convergence constants k_1 and k_2. In the spatial affine grid context the search space and neighbourhood size are already dependant on the grid region dimensions. When the convergence parameters are chosen to satisfy approximately $k_2 = 10k_1$ the algorithm reaches equilibrium in s small number of iterations. Taking these considerations, the parameter values were: $d_{max} = 0.1$, $w_{AB} = 7$, $w_{DF} = 2$, $d_{dif} = 0.05$, $k_1 = 0.2, k_2 = 2$, $V = 0.2$, $N_{it} = 3$. Figure 8 illustrates the matching of

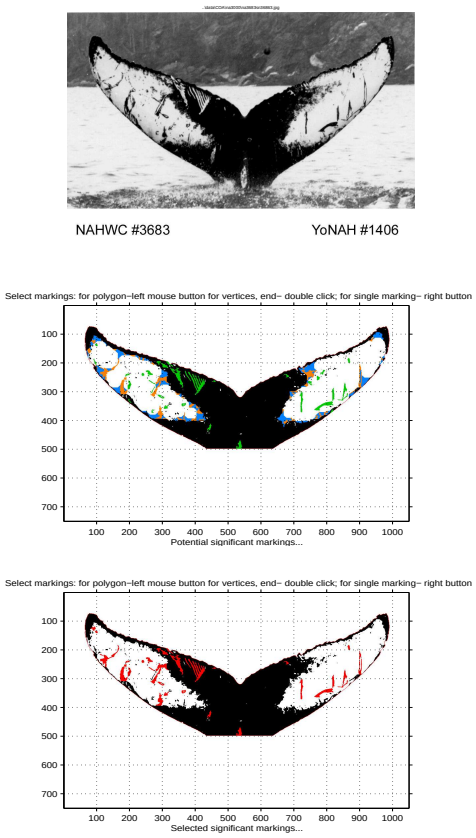

Figure 6: Original humpback fluke image (top) (also in Fig. 1), the salient patterns detected (middle) and the selected significant patterns (bottom).

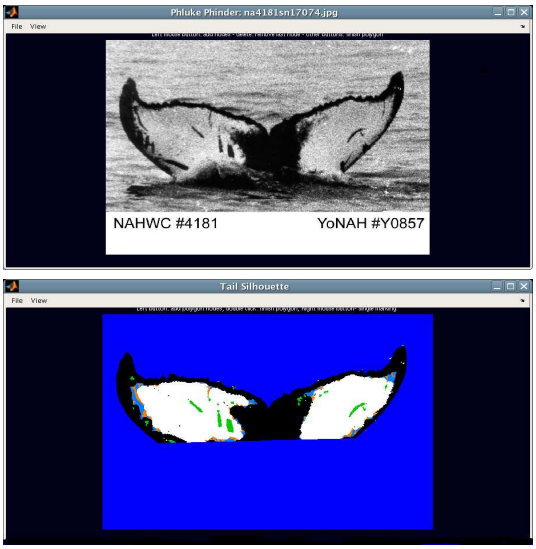

Figure 7: Original humpback fluke image (top), and the the salient patterns detected (bottom)

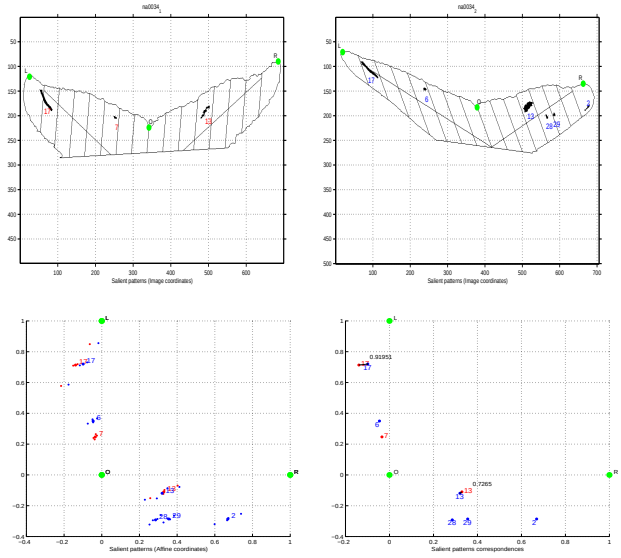

Figure 8: Top: segmented flukes contours of images of the same individual: 'n34_1'(left) and 'n34_2' (right) and their salient patterns. Bottom: the patterns representation by affine coordinates (left) and the patterns matching (right).

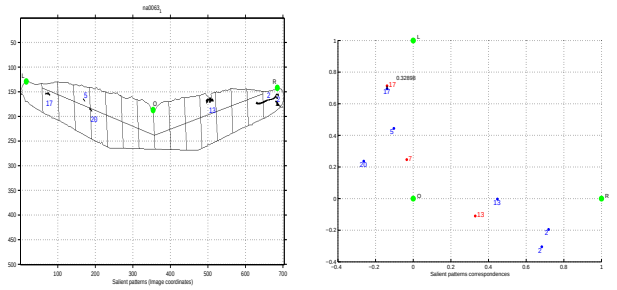

Figure 9: Segmented fluke contour for image 'n63_1' (left). The patterns matching with 'n34_1' (right).

a pair of images of the same individual. Next to each pattern, the affine grid region number of the pattern centroid is displayed. Note that the matching patterns are in regions 17 and 13. The number of patterns for different images of the same individual depend on the quality of the segmentation, hence on the photographic quality and occlusion by the sea. The manual selection of patterns subset may also introduce such difference. In the point-set representation sub-figure for each pattern, the centroid (larger dot) and the 4 extreme points are displayed. The matching sub-figure shows all detected pattern correspondences and their weights. Note that the algorithm has detected the 2 correspondences assigning them high weights.

Figure 9 shows another individual, the patterns and the matching to the first image of the pair from Figure 8. The correspondence in regions 17 was assigned a low weight.

Table 1 gives the SP similarity score (19) which

Table 1: Salient pattern similarity ($w_l = 0.6$) for 'n34_1' with its true match and with a mismatch.

Image	Σ_w	Σ_{w_l}	s_M	s_I	s_{IL}	s_{SP}
'n34_2'	1.646	1.646	0.25	0.7	0.7	**5.432**
'n63_1'	0.329	0	0.25	0.5	0.393	**0.376**

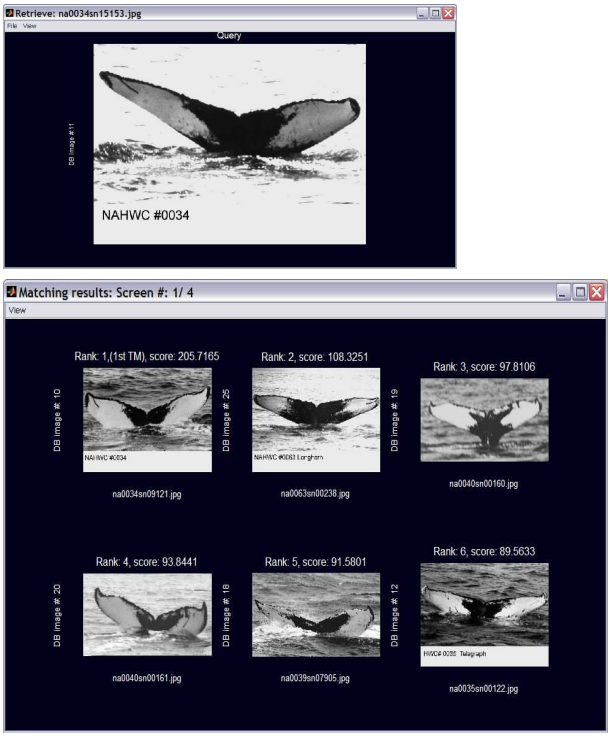

Figure 10: Screen-shots of the retrieval system. Top: Query. Bottom: Retrieval results for the CB method. The true match is ranked first.

distinguishes well between a match and a mismatch.

5.3 Retrieval

The system presents the retrieval result as a ranked list of potential matches for visual inspection by an expert (6 images per screen) with the similarity scores. Figure 10 gives an example of the output of the retrieval system for the CB method. The method based on black and white ratios orders all tails of types compatible with the query type T_q, which for the example are 241 images presented on 41 screens. The true match is ranked at position 20 on the forth screen. The SP method comparing only retrieved markings ranks the true match second. The combined method perform SP matching only among the top 21 images (on four screens) given by the BW method. The true match is ranked first.

Figure 11 gives an example of a more difficult case, due to the quality of the query's true match. The BW method ranks the true match at position 19, while the SP method assigns rank 6. Combining the two methods obtains the best result- rank 1.

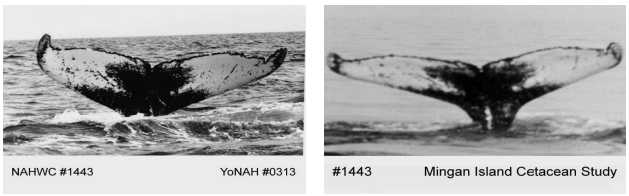

Figure 11: Query and its true match of poor quality.

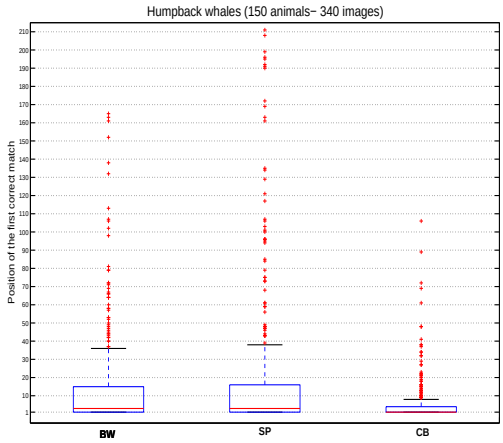

Figure 12: Box-plots exhibiting identification performances for BW, SP and CB methods applied to the whole DB.

The overall performance of the retrieval methods is illustrated in Figure 12. Box-plots have been used to show the position of the first correct match for various percentiles of the test DB. A box-plot has horizontal lines at the lower quartile, median (shown in red) and upper quartile levels. It can be seen that the CB outperformed both the BW and the SP methods with the excellent median level rank value of 1. On the other hand, the SP and CB methods are computationally more intensive than the BW method. The BW method took only 12 seconds for all images to be matched against the rest of the DB, while the elapsed times for SP and CB were 105 mins. (average of about 20 s. per image) and about 60 mins. (average of 10 s. per image) respectively.

The overall identification results are very good as summarised also in Table 2. For the CB method about 60% of the time the first correct match was ranked first and for 80% of the time the user could locate the true match on the first screen (5 images), which is a considerable reduction in the search time. These results are superior compared to similar studies. The best rank obtained by the curve-matching technique in [5] for the median level is 9, while in [7], rank 5 was reported for similar experimental settings for a grey whale DB. Our method has the advantage of avoiding the high order moment invariants used in [7] on the cost of a larger involvement of the human user.

Table 2: Percentage of images whose first true match is ranked amongst the top k.

Method	$k = 50$	$k = 20$	$k = 10$	$k = 5$	$k = 1$
BW	90.9	79.4	70.6	60	36.2
SP	84.1	73.8	65.8	55.8	36.2
CB	96.7	92.2	84.7	77.9	60.5

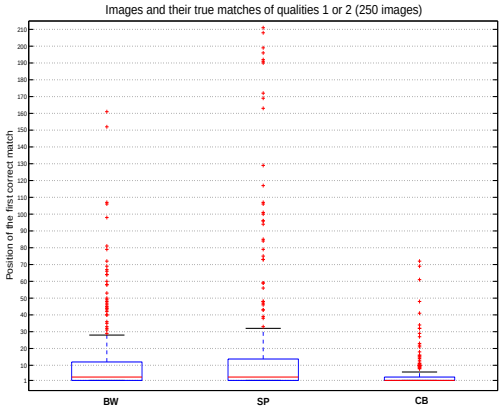

Figure 13: Box-plots exhibiting identification performances for BW, SP and CB methods for part of the DB. The images and their true matches have excellent to good recognition and photographic qualities.

Figures 13 and 14 illustrate the effect of the image quality on the retrieval performance. The former summarises the performance of the three methods for images and their true matches which have qualities of 1 and 2, while the latter - for qualities 2 and 3. Again the CB method shows best robustness to image quality. Comparing Figures 12 and 13, it can be concluded that the performance only very slightly deteriorates when poor quality images are present in the DB. Figure 14 shows that even if only poor quality images are considered, 75% of the time the CB method ranks the first true match within the top 9 images.

6 Conclusion

In this paper we present our computer-assisted photo-identification system for humpback whales as an example of a content-based image retrieval problem. In an effort to simulate the human perception of saliency and similarity, we propose a novel binary salient pattern detector and a point-matching strategy. They are general purpose tools and can be applied for photo-identification of other species and in different application contexts. The retrieval system combines features based on simple black and white ratios with the detected and matched salient patterns representations. The performance plots clearly show the increased robustness on a speed which allow the system to be

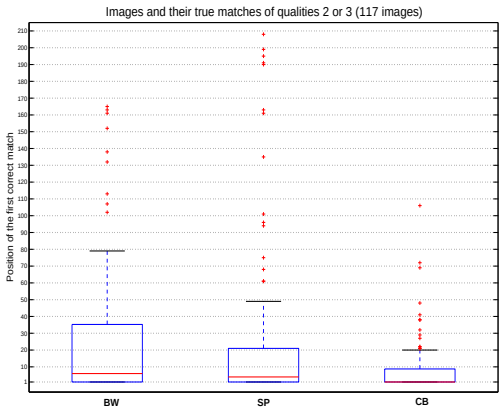

Figure 14: Box-plots exhibiting identification performances for BW, SP and CB methods for part of the DB. The images and their true matches have moderate to poor recognition and photographic qualities.

used by the marine biologists in their identification studies. The search time compared to manual photo-identification is considerably reduced. The system has been welcomed favourably by experts.

Acknowledgements

This work has been supported by the Dutch science foundation project " Computer-assisted photo-identification of cetaceans" (NWO 613.002.056) and the EC-sponsored Network of Excellence MUSCLE (www.muscle-noe.net). The authors thank Judy Allen from College of the Atlantic, Maine, US for for providing the image database and for her expert advices and help with the segmentation.

References

[1] B. Araabi. *Syntactic/Semantic Curve-Matching and Photo-Identification of Dolphins and Whales.* PhD thesis, Texas A&M University, 2001.

[2] S. Barnard and T. Thompson. Disparity analysis of images. *IEEE Transactions on PAMI*, 2(4):333–340, 1980.

[3] B. Araabi et al. Generalization of Dorsal Fin Ratio for Dolphin Photo-Identification. *Annals of Biomedical Engineering*, 28(10):1269–1279, 2000.

[4] B. Araabi et al. Locating an Affine/Projective Invariant Identifier Patch on an Image. In *proceedings of the 5^{th} IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 121–125, 2002.

[5] C. Gope et al. An affine invariant curve matching method for photo -identification of marine mammals. *Pattern Recognition*, 38:125–132, 2005.

[6] G. Hillman 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*, Dec. 2003.

[7] N. Kehtarnavaz et al. Photo-identification of humpback and gray whales using affine moment invariants. In *Proceedings of 13th Scandinavian Conference on Image Analysis*, pages 109–116, 2003.

[8] S. Mizroch, J. Beard, and M. Lynde. Computer Assisted Photo- Identification of Humpback Whales. In P. Hammond, S. Mizroch, and G. Donovan, editors, *Individual Recognition of Cetaceans*, pages 63 – 70. International Whaling Commission, Cambridge, 1990.

[9] E. Ranguelova, M. Huiskes, and E. Pauwels. Towards computer-assisted photo-identification of humpback whales. In *Proceedings of ICIP04*, pages 1727 – 1730, October 2004.

[10] P. Soille. *Morphological Image Analysis.* Springer, 2003.

[11] B. Super. Fast correspondence-based system for shape retrieval. *Pattern Recognition Letters*, 25:271 – 225, 2004.