- Invited paper -

Wild Cetacea Identification using Image Metadata

Débora Pollicelli[1,2], Mariano Coscarella[1,3], and Claudio Delrieux[4]

[1] *CESIMAR-CONICET, Centro para el Estudio de Sistemas Marinos, Consejo Nacional de Investigaciones Científicas y Técnicas, CCT CENPAT, Bv. Almirante Brown 2915, 9120, Puerto Madryn, Chubut, Argentina*
[2] *LINVI, Departamento de Informática, Facultad de Ingeniería, UNPSJB, Bv. Almirante Brown 3051, 9120, Puerto Madryn, Chubut, Argentina*
[3] *Departamento de Biología General, Facultad de Ciencias Nauturales, UNPSJB, 9120, Puerto Madryn, Chubut, Argentina*
{deborapollicelli, mcoscarella}@gmail.com
[4] *Laboratorio de Ciencias de las Imágenes, Departamento de Ingeniería Eléctrica y Computadoras, Universidad Nacional del Sur y CONICET, 8000 Bahía Blanca, Argentina – cad@uns.edu.ar*

Abstract

Identification of individuals in marine species, especially in Cetacea, is a critical task in several biological and ecological endeavours. Most of the times this is performed through human-assisted matching within a set of pictures taken in different campaigns during several years and spread around wide geographical regions. This requires that the scientists perform laborious tasks in searching through archives of images, demanding a significant cognitive burden which may be prone to intra- and interobserver operational errors. On the other hand, additional available information, in particular the metadata associated to every image, is not fully taken advantage of. The present work presents the result of applying machine learning techniques over the metadata of archives of images as an aid in the process of manual identification. The method was tested on a database containing several pictures of 223 different Commerson's dolphins (*Cephalorhynchus commersoni*) taken over a span of seven years. A supervised classifier trained with identifications made by the researchers was able to identify correctly above 90% of the individuals on the test set using only the metadata present in the image files. This reduces significantly the number of images to be manually compared, and therefore the time and errors associated with the assisted identification process.

Keywords: machine learning, photo-identification, cetaceans, Commerson's dolphins

1 Marine Mammal Individual Identification

In Biology, Ecology, and other sciences, the ability to recognize individuals allows researchers to obtain relevant information that is crucial for several scientific purposes, including population parameters estimation such as size, fertility, survival and mortality rates, home ranges and movements, etc. [1, 2]. These parameters are usually derived or inferred from the implementation of *capture-recapture* models. Capture-recapture models are based on the possibility of identifying a specific animal (individual) from one sampling occasion to another, considering the first time the animal was photographically registered as a "capture" and the subsequent times as "recaptures" [3]. Since the 1970s, researchers relied on natural marks or other visual features to identify animals with non-invasive means. This picture-based identification technique was developed for cetaceans or other large marine fauna, mainly because handling other recognition means (f.e., attaching straps or belts to the individuals) is expensive, difficult and invasive, being impractical as an identification mean in the field. On the other hand, taking pictures (captures or recaptures) is relatively inexpensive and less difficult, providing reliable information on which were the individuals present at a given place and time, with the obvious disadvantage of depending on futher recaptures of the individual and a proper identification in the picture archives.

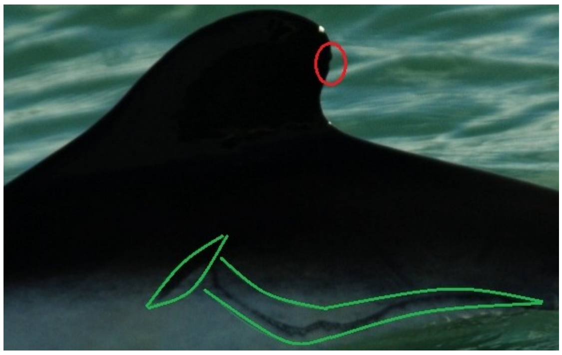
Recognition of an individual cetacean in pictures is usually performed using different features. For example, southern right whales (*Eubalaena australis*) may be identified using the callosities patterns located in the upper part of the whales' head. Recognition of notches and scars in the edge trail of the fluke is common for sperm whales (*Physeter macrocephalus*) and humpback whales (*Megaptra novaeangliae*), and the shape and notches on the dorsal fin is used in the identification of the killer whales (*Orcinus orca*) or the bottle-nose dolphins (*Tursiops sp.*) [4, 5, 6].

As mentioned above, human-assisted recognition of dolphins and whales using pictures is a difficult and time consuming task. For this reason, some software products are available to assist researchers on this task, like DARWIN [7, 8]. However, these products are neither effective in all cetacean species, nor useful among species in which the same type of feature is used to produce the individual recognition. In particular, they require a quite accurate supervised landmarking, including identifying the tip of the fin, and the position of the notches to be able to compensate for the perspective distortion in taking the picture, an unrealistic requirement most of the times [7].

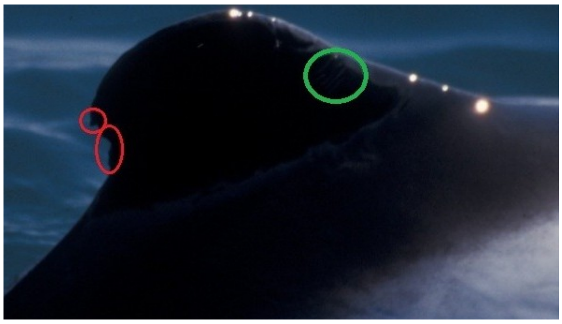
A major source of false negatives in individual identification in these systems is produced due to the unsuccessful application of 3D correction before matching a given record with previously identified individuals. This is a critical issue, because pixel-based matching (for instance, using Euclid distance) is not robust under landmark positioning differences, which are almost certain to occur due to intra and intersubjective appreciation errors. For this reason, the success of landmarking over images as an identification means is tied to the operators' ability to produce accurate lanmarkings consistently. For this and other reasons, according to Stewman [9], landmarking is not entirely reliable, and additional information is required during the record registration to optimize further identifications.

Even more difficult is identification in a *Genus* of southern hemisphere dolphins that have some species with rounded dorsal fins, because it is not possible to pinpoint landmarks. The *Cephalorhynchus* species, and particularly *C. commersonii*, require for their individual recognition to rely on the traditional method in which the operator is trained to find matches manually. The notches in the trailing edge of the dorsal fin, and also color variation patterns, are used for identification. The notches are visible at different angles, and therefore are more likely to be useful in photo-identification. In contraposition, other kind of scars and abnormalities in the coloration patterns are used as ancillary features, since generally they allow to identify the animal from only one side.

So far, no reference in the literature proposes the use of the metadata associated to the imagery as a filtering means to lighten and speed up the matching task. The purpose of this presentation is to show the preliminary results of a research line aimed to automatize marine mammal photo-identification. Apart from image-based techniques as the ones mentioned above, the ancillary information present in the photographic database is not taken advantage of. In a series of studies carried out in the Patagonian coast, a database of individually recognized Commerson's dolphins had been kept in the LAMAMA-CECIMAR-CONICET Institute [10]. The information accrued includes not only pictures but also a series of dolphins' descriptors [11] (see Fig. 1). In this work we show how this information can be used in the context of automated recognition of individuals, achieving an identification accuracy above 90% employing only the images' metadata. This alleviates the cognitive burden of the researchers in applying the capture-recapture model, and shows that metadata combined with image-based techniques may derive new automated identification products that go beyond the state-of-the-art in marine mammal photo-identification.



(a) Subtle notch and large auxiliary mark visible only from the left side.



(b) Multiple visible notches and subtle auxiliary mark on the right side.

Figure 1: Individually identified Commerson's dolphins in the LAMAMA-CESIMAR-CONICET data base. The red areas show notches in the trail of the dorsal fin. These are considered primary marks. The (often more subtle) auxiliary marks are shown in green. Primary marks are feasible to be recognized from both sides of the animal, while auxiliary marks generally are visible from only one side.

2 Materials and methods

The preparation and process data were done in four stages following the methodology proposed by Ferrary [12] and Witten [13] for data mining procedures:

1. Define the goals and the information sources, and collect the data.

2. Analyze and preprocess the data.

3. Build and train models.

4. Perform validation tests.

In what follows of this Section we describe each of these steps (see also [14]).

2.1 Data collection and analysis

As stated above assisting in the identification of the dolphins can significantly reduce the operator's time, by reducing the number of photographs to browse. We propose the use of a classification model that aids in the matching process using patterns present in the pictures' metadata. Also, we aim to determine how similar are the marks of certain identified animals. The information is persisted in 869 $MSAccess^{TM}$ database records that hold the data and pictures of a population of Commerson's dolphins, spanning along seven years, that have a total of 223 identified dolphins. These records, together with additional metadata used for photo-identification are used as instances (examples). From these instances we preselected only the specific attributes that may be relevant in the photo-identification task (see next subsection). Then the data was migrated to $MSExcel^{TM}$, where data wrangling procedures were applied for data extraction and cleansing. Finally, numerical values were asigned to nominal attributes, and to text attributes indicating ordinal values.

2.2 Attribute selection and data cleansing

A set of attributes that *a priori* hold significant information that could assist the photo-identification task were initially preselected to train the classifier:

- **Side**. The side of the animal where the picture was taken ("right" or "left"). The scars and amount of coloration attributes clearly depend on the this attribute for a given individual.

- **Quality**. A quality index between 0 and 3 is assigned, related to image quality features including brightness and contrast, fin correctly focused, fin vertically aligned, and presence of water waves or drops obscuring the fin.

- **Distinctiveness**. A distinctiveness index between 0 and 3 is assigned given by the intrinsic features of the fin, including how visible or distinguishable are the notches and marks in the edge of the fin.

- **Scars**: A numerical quantity that represents the amount of recognizable scars observed in the picture. This attribute is related to side, quality and distinctiveness.

- **Coloration**. A numerical quantity that represents the amount of recognizable abnormal coloration spots observed in the picture. Also related to side, quality and distinctiveness.

- **Zones**. Specific areas in which the notches and marks may appear in the dorsal fin are designated with numbers 1 to 7 (see Fig. 2). This attribute takes a "true" value if the individual have notches or marks in this zone, and "false" otherwise.

- **Notches**. A numerical quantity that represents the amount of recognizable notches or marks in the edges of the fin. Not necessarily equal to the sum of all "true" values in the zone attributes since a notch may involve more than one zone, and also in a zone more than one notch may be located.

- **Catalog Number**. A unique id number for each identified animal.

- **"Big/large/extended", "Medium", "Small/little"**. These attributes describe the amount of marks with this size feature.

- **"Little bit/mild/imperceptible", "Triangular", "Rounded", "Salience"**. These attributes describe the amount of marks with this shape feature.

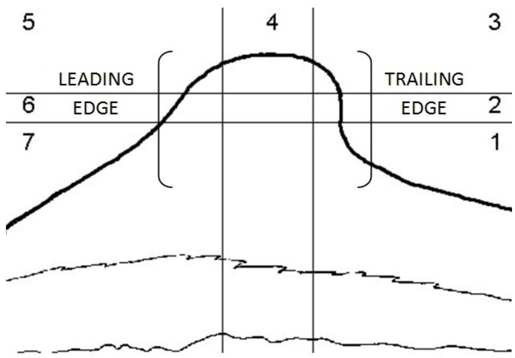


Figure 2: The seven zones in the dorsal fin.

The latter two attributes had to be carefully checked in order to be meaningful. Natural language attributes are prone to spelling and wording errors, and therefore disambiguation was required. Records which had incomplete information were discarded. Also, only the records that were originally used for photo-identification were considered.

3 Results

3.1 Attribute selection and classification methods

Four different supervised classification algorithms, each pertaining to a different classification method,

Table 1: Accuracy (in %) of the three different classifiers with different subsets of attributes.

Dataset Name	Naive Bayes	KStar	trees J48
full set	46.83	49.13	43.41
full set - Z7	67.73	68.35	66.02
full set - Z7 - Q	66.68	68.35	65.40
full set - Z7 - Q - R	66.62	67.43	64.57
full set - Z7 - Q - R - T	66.84	67.92	64.76

were used:

- Neural networks: **Multilayer Perceptron**
- Bayesian classificators: **NaïveBayes**
- Decision trees: **J48**
- K-nearest neighbor algorithm: **KStar**

To avoid overfitting, the attributes were selected using `Info Gain Attribute Eval`, `Gain Ratio Attribute Eval` and `Chi Squared Attribute Eval` in conjunction with the Ranker search method, that ranked all attributes by their individual evaluations [15, 16]. In all trials, the results showed that the attributes *Rounded (R)*, *Triangular (T)*, *Zone 7 (Z7)* and *Quality (Q)* were mostly weighless and therefore were discarded. Also `Cfs Subset Eval` combined with Best First, showed the same behavior for attributes *Little bit/mild/imperceptible (L)* and *Zone 4 (Z4)*. Removing some attributes we improved the accuracy of the classifier, with respect to the full set of attributes. In Table 1 the obtained accuracy of gradually subtracting these attributes can be appreciated.

3.2 Model construction and validation

Once the dataset (instances and attributes) was cleansed and filtered, a standard cross-validation procedure training was first performed. The dataset included 869 instances of 223 individuals. It is worth to note that the amount of "recaptures" of each individual is very uneven, ranging from 1 in most cases up to 24 in one case. Thus, the classes are unbalanced and therefore special consideration must be taken during the model construction to avoid biasing the classifier[17]. In our case, we splitted the dataset into three groups, according to the amount of recaptures of each individual in ranges from 1 to 4, 5 to 12, and more than 12. In the first group, the amount of instances per individual is too low to achieve a significant accuracy. On the other hand, in the third group the amount of individuals is too low (only eight), with a large amount of recaptures. For this reason, excluding these examples would avoid unbalancing the classes during learning without severely limiting the amount of individuals identified. We additionally used a fourth group that comprised individuals with 5 or more recaptures. Thus, in Table 2, the results of classification tests aiming to compare the results variation between the the 5 to 12 and more than 5 recaptures are shown. The subsets that were considered to be adequate were:

- Subset between 5 to 12 recaptures with a total of 373 instances of 54 individuals.
- Subset with greater than or equal to 5 recaptures with a total of 515 instances of 62 individuals.

To test whether the training set is statistically meaningful, the ZeroR classifier was applied to check the accuracy of the majority class. The obtained result of 2.4862% correctly classified instances was well above the 1.8% ($frac154$) expected by pure chance. For model construction, each subset was split into training set and validation set. These split were made in two ways 90%-10% and 97%-3%. In all cases, the training was performed using the cross-validation technique with a k=10 (folds) with the same four learning methods. The accuracies of the four methods are shown in Table 3.

For both dataset and in both ways split training set the accuracies were approximately the same, varying between 61.12% and 68.31%, using differents classifiers.

3.3 Validation

Once the classifiers were trained with the filtered training set, we tested them with the instances in the validation set. This is the final intended use of the system, since these examples act as if they were new captures of already captured animals (see Table 4). In this situation, the accuracy varied between 56.86% and a 90% for different classifiers. Nevertheless, the smaller the validation set, the higher the classification percentage. For the 3% validation set the figures ranged from 72.72% to 90%, while when using the 10% validation set these values dropped to a range between 56.86% and 72.97%. There is no clear pattern on whether the balanced group (5 to 12 recaptures) outperform the more than 5 recaptures set. When considering the 10% validation partition the best results are

Table 2: Accuracy (in %) obtained with the four different classifiers with the complete dataset and the balanced subsets.

Dataset	**Naive Bayes**	**K Star**	**J48**	**Multilayer Perceptron**
Complete	46.69	49.14	43.34	41.64
5 to 12	67.73	68.35	66.02	63.47
≥ 5	66.99	69.51	66.79	63.88

Table 3: Accurracy (in %) of the four classifiers with the two filtered training set.

Training split	**Dataset**	**Naive Bayes**	**KStar**	**J48**	**Multilayer Perceptron**
90%	**5 to 12**	66.66	65.77	63.98	62.5
	$\geq$ **5**	67.02	68.31	66.16	61.42
97%	**5 to 12**	66.02	67.4	62.98	62.7
	$\geq$ **5**	67.13	67.53	66.73	64.12

obtained for the 5 to 12 recaptures set, while for the 3%validation partition, the best results are obtained for the more than 5 recaptures set. Anyway, the smaller the validation set, the better the results.

4 Discussion and conclusion

We presented the result of applying machine learning techniques over the metadata of archives of 869 pictures taken of 223 different Commerson's dolphins images, as an aid in the process of manual identification of individuals. The metadata consisted of a set of manually taken annotations, one record per picture, that described different aspects of the animal's fin and surrounding appearance, together with ancillary information regarding the place and time where the picture was taken. The metadata was arranged as a set of attributes, and incomplete or incorrect records were filtered out. Attributes were further curated for schema conformance, mapping annotated values to numerical or ordinal categories adequate for the automated learning process. Finally, superfluous or noisy attributes were filtered out.

Preliminary results showed that animals with few pictures (below 5) were almost impossible to identify with only this metadata. Therefore the learning algorithm was focused only on animals with greater than or equal to 5 recaptures records for each individual. A supervised classifier was trained with the identifications provided by the biologists. These results show that the system may be quite helpful in the task of reducing the supervised time and effort of identification of new pictures, at least if there is a representative amount of priorly taken pictures of the same individual.

As previously suspected, the results shown that the smaller the validation set the higher the accuracy, and hence the correct identification of the individuals. This gives an insight on the way this algorithm should be used for this particular purpose. As new individuals are gathered in each photographic session [11], the number of photographs to be analyzed after will be small (never exceeding 5% of the database photographs). Hence, the a validation set of 3% is roughly reasonable to be tested. Nevertheless, before each new identification session, the algorithm should be re-trained, in order to include the new photographed individuals. The problem associated with the seldom captured animals (less than four captures) is pending to be resolved.

Current work around this project is focused on enhancing the accuracy on seldomly recaptured animals. Using metadata only, the semantics of the manual annotation can be further mined using text mining to deliver a more fine-grained set of nominal attributes regarding the description of the shapes and coloration of notches and marks, using a convenient thesaurus. Also, we are currently working on image analytics, using first HaarCascade descriptors for ROI automatic detection (mainly of the fin in the pictures) and then morphometric descriptors to obtain an additional feature vector that combined with the available metadata may achieve better identification performance. Finally, we are considering other analytic features of the global population of captured animals. For instance, performing spatio-temporal analysis of capture-recapture patterns may reveal trends that may further aid in the automated identification process.

References

[1] K. H. Pollock, J. D. Nichols, C. Brownie, and J. E. Hines, *Statistical inference for capture-recapture experiments*, vol. 107 of *Wildlife Monographs*. Blacksburg: The Wildlife Society Inc., 1990.

[2] H. Whitehead, "Mark-recapture estimates

Table 4: Validation results (accuracies in %)

Validation split	Dataset	Naive Bayes	KStar	J48	Multilayer Perceptron
10%	**5 to 12**	72.97	75.67	70.27	70.27
	$\geq$ **5**	68.62	62.74	64.70	56.86
3%	**5 to 12**	81.81	81.81	90	72.72
	$\geq$ **5**	87.5	87.5	81.25	81.25

with emigration and re-immigration," *Biometrics*, vol. 46, pp. 473–479, 1990.

[3] R. M. Cormack, *Models for capture-recapture*, pp. 217–255. Maryland: International Cooperative Publishing House, 1979.

[4] M.Bigg, "An assessment of killer whale (orcinus orca) stocks of vancouver island, british columbia," report, International Whaling Commission, 1982.

[5] P. S. Hammond, S. A. Mizroch, and G. P.Donovan, *Individual recognition of cetaceans: Use of photo-identification and other techniques to estimate population parameters*, vol. 12 of *International Whaling Commission Special Issue Series*. Cambridge: International Whaling Commission, 1990.

[6] B. WÃijrsig and M. WÃijrsig, "The photographic determination of group size, composition and stability of coastal porpoises (tursiops truncatus)," *Science*, vol. 198, pp. 755–756, 1977.

[7] E. College, "Darwin." `http://darwin.ekerd.edu/`.

[8] T. A. University, "Fin scan."

[9] J. Stewman, K. Debure, S. Hale, and A. Russell, "Iterative 3-d pose correction and content-based image retrieval for dorsal fin recognition," in *International Conference Image Analysis and Recognition*, pp. 648–660, Springer, 2006.

[10] M. A. Coscarella, S. Gowans, S. N. Pedraza, and E. A. Crespo, "Influence of body size and ranging patterns on delphinid sociality: Associations among commerson's dolphins," *Journal of Mammalogy*, vol. 92, no. 3, pp. 544–551, 2011.

[11] M. A. Coscarella, *Ecología, comportamiento y evaluación del impacto de embarcaciones sobre manadas de tonina overa Cephalorhynchus commersonii en Bahía Engaño, Chubut.* Ph.d., 2005.

[12] M. J. R. Q. C. Ferri Ramírez Alhambra, J. H. Orallo, *Introducción a la minería de datos.* Pearson, 2004.

[13] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques.* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 3rd ed., 2011.

[14] M. Hall, E. Frank, G. Holmes, B. Pfahringer, and I. H. Witten, "The WEKA data mining software: An update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.

[15] M. L. G. at the University of Waikato, "Weka 3: Data mining software in java." `http://www.cs.waikato.ac.nz/ml/weka/`.

[16] R. R. Bouckaert, E. Frank, M. Hall, R. Kirkby, P. Reutemann, A. Seewald, and D. Scuse, *WEKA Manual for Version 3-6-12*, 2014.

[17] O. L. González, J. F. M. Trinidad, and M. G. Borroto, "Clasificadores supervisados basados en patrones emergentes para bases de datos con clases desbalanceadas," Tech. Rep. CCC-14-004.14, Coordinación de Ciencias Computacionales. INAOE, Luis Enrique Erro 1. Sta. Ma. Tonantzintla, 72840, Puebla, México., 2014.