Marine Life Observation using Classification Algorithms on Ocean Surface Photographs

Frédéric Maussang[1], Liyun Guelton[1], René Garello[1], and Alexis Chevallier[2]
[1]Telecom Bretagne, UeB, CNRS Lab-STICC
Technopôle Brest-Iroise, CS 83818, 29238 Brest Cedex 3, France
Email: {frederic.maussang; liyun.guelton; rene.garello}@telecom-bretagne.eu
[2]IN VIVO Environnement
ZA La Grande Halte
29940 La Forêt-Fouesnant, France
Email: alexis.chavallier@invivo-environnement.com

***Abstract*—The growth of marine renewable energy and marine protected areas in France leads to an increased need for animal population knowledge at sea. Offshore energy generator projects (wind turbines for example) must obey these regulations and show their harmlessness to the environment, particularly to the wildlife and to protected species, which are vulnerable and threatened. The work presented in this paper is part of a project whose goal is to propose a system of sea surface survey by plane, using numerical HD photography. Data will be recorded in daylight with a flight plan allowing to cover all of the sea surface to be studied. This paper introduces a solution to use these images to detect and classify the objects present on the sea surface, near this surface or in the air column under the plane: birds, marine mammals, turtles and sharks (near the surface), ships, other mankind objects (waste).**

I. Introduction

The growth of marine renewable energy and marine protected areas in France leads to an increased need for animal population knowledge at sea. Moreover, new legal dispositions require to study the potential impacts of all major projects at sea, including impacts on the environment and biota. Offshore energy generator projects (wind turbines for example) must obey these regulations and show their harmlessness to the environment, particularly to the wildlife and to protected species, which are vulnerable and threatened. However, the surfaces to observe are very large and today there is no commercial autonomous monitoring system: birds and marine mammals are observed and counted manually from a plane or a ship [1].

The work presented in this paper is part of a project whose goal is to propose a system of sea surface survey by plane, using numerical HD photography.

This paper introduces a solution to use these images to detect and classify the objects present on the sea surface, near this surface or in the air column under the plane: birds, marine mammals, turtles and sharks (near the surface), ships, other mankind objects (waste).

This paper is then divided into two main parts. A first section will discuss the interest of airborne photography in marine animal population estimation. A second one will be on the proposed detection approach based on a learning process taking into account the variety of possible targets and their motions. Finally, a short conclusion will be done and some perspectives open.

II. Airborne photography

A. Interest of airborne photography

Today, birds and marine mammals populations near the French coast are estimated visually by humans from an harbor, a boat or a plane. Thanks to the expertise of the persons implied in these campaigns, knowing the physical characteristics and behaviors of the researched species, these estimations are relatively good.

But these activities are expensive in human resource and and the results are highly dependent from the weather, sea surface phenomenons and reflectivity.

That is why the interest of sea surface imagery is real. Moreover, as Terletzky *et al.* said [2], this offers the following advantages: data can be used a long time after the registration, allowing to perform efficient processing techniques on it, to extract information invisible from a human (e.g. spectral), some images can be eliminated due to a bad visibility, the presence of birds to close to the receiver. Elements on the images (coast, rocks, buildings. . .) can also be used as references to localize each image, and then avoid to count several times the same targets.

B. Images capture

Concerning our project, data will be recorded in daylight with a flight plan allowing to cover all of the sea surface to be studied. The images obtained will be in RGB (Red, Green, Blue), representing approximately 400m long and 300m large, with a resolution of a little more than 3cm on the surface (Fig. 1). The flight will be done with a high of 1000 feet approximately and a speed of 220 km per hour. Its plan is built in order to have both the best covering of the studied area and the lowest overlap between images, to avoid multiple views of the same targets.

An example of airborne sea surface image, similar to the images we will obtain thanks to the system, is presented in Fig. 2, presenting some targets (birds). We note the crucial

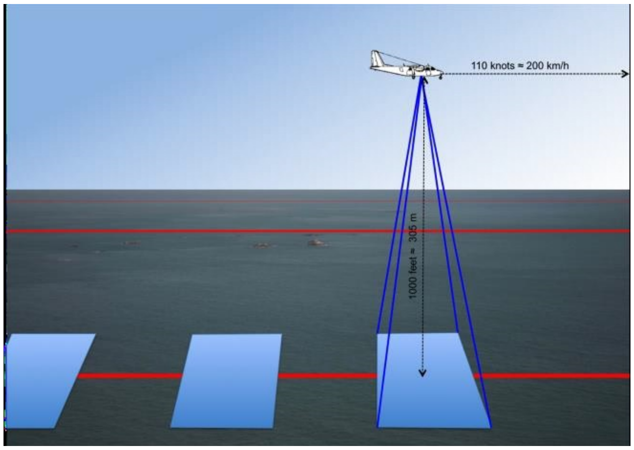

Fig. 1: Marine environment marine survey using airborne photography system (images of 400m x 300m)

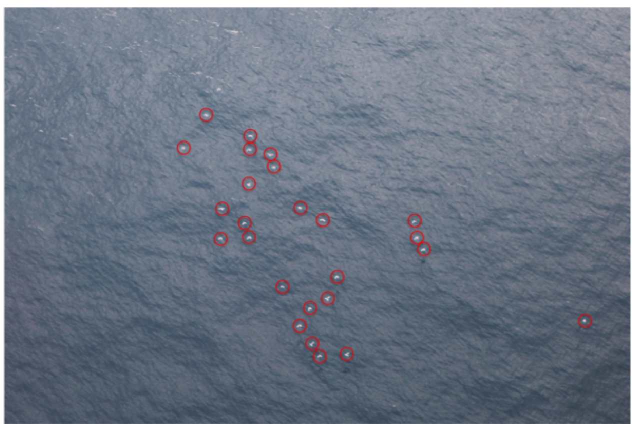

Fig. 2: Example of images: target circles in red represent the object we want to detect and classify. See Fig. 3 to have examples of zoomed images

place of the resolution in detection and recognition of these targets. Indeed, the detection, and moreover the recognition, need to have a sufficient number of pixels to be efficient. An other aspects is in the distance between the receiver, in the plane, and the target: its occupation will vary with this distance, with particular cases of too large occupations leading to the elimination of the corresponding images in the classification process. These remarks will be taken into account in the choice and the limitations of methods described following.

To be able to identify automatically the marine animals and birds, an image database containing their thumbnails from the full size image database is needed [3]. In each thumbnail, there is only one object (animal, bird, rock, sea texture or other, see Fig. 3). These thumbnails will be already classified, but manually. Then they can be used to learn a classification model allowing to map the thumbnails from a new input image to their corresponding classes.

(a) Turtle

(b) Bird

(c) Rock

(d) Sea texture

Fig. 3: Four object classes: marine animal (a), marine bird (b), rock (c), and sea surface (d); ©Kelonia/Sensefly

III. Learning-based object detection

Humans can recognize a variety of objects in images with little effort, despite that the images of the same object may vary a lot in color, size and scale, or even when they are translated or rotated. However, this task is still a challenge for computer vision systems where object identification algorithms rely on model-based matching, pattern recognition or learning-based recognition. These algorithms involve often appearance-based or feature-based techniques including gradients, edges, Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT) and so on.

In this study, marine animals, birds or clutter issues have a variety of shapes, poses and textures. Moreover, detection objects are deformable, submerged, and are of low resolution. It seems too complex to apply a rigid model which looks for exact matches in the input with pre-existing models or patterns, such as the approaches explored in [3] for the identification of white sharks in aerial image sequences. In this study, we investigate learning-based models.

A. Supervised classification modelling

The learning based problem is formulated as:

$$y = f(x) \tag{1}$$

where $x \in \mathrm{X}$ is input instance and $y \in \mathrm{Y}$ is output label. Our goal is to use a training set $D = \{(x_i, y_i) \,|\, x_i \in \mathrm{R}^p, y_i \in \{1, \cdots, k\}\}_{i=1}^n$, where k is number of classes, to produce a mapping function $f' : \mathrm{X} \to \mathrm{Y}$ as close as possible to the real mapping function f.

To retrieve the optimal mapping function f', in general, a loss function is defined. It assigns a specific value to mistake resulting from producing an incorrect label. The goal is finally to minimize the expected lost. This general framework involves two key elements : the feature extraction x issued from aerial images and the definition of classification model f.

B. Feature extraction

From an initial set of measured data instances, feature extraction algorithms derive values (features) intended to be more informative, non redundant. Feature extraction is an essential pre-processing step to pattern recognition and machine learning problems [4], as it facilitates the subsequent learning and generalization steps. Before feature extraction step, a phase of de-noising, smoothing, or sharpening is usually applied to improve signal-to-noise ratio. The Fourier transform and wavelet transforms are popular methods for this. When the dimensionality of the data is very high, some techniques such as Principal Component Analysis (PCA) might be used to project the data into a lower dimensional space in keeping as much information as possible.

A very large number of feature detectors has been developed, such as edges [5], corners and blobs detection/regions of interest [6], HOG [7], SIFT [8] and so on. Once features have been detected, a local image patch around the feature can be extracted. These convolutional-based methods to extract local features use hand-crafted kernels or syntactic and structural methods [4]. However, manual feature extraction is a tedious process requiring extensive domain and vision expertise [9]. Higher level algorithm may be used to guide the feature detection stage, so that both local and global properties as well as spatial relationships can be taken into consideration. This motivates the design of efficient feature learning techniques, which will be discussed in section III-D.

C. Classification models

The definition of classification model f turns out the same as the decision to be made concerning what kind of classifier to apply to feature vector. Popular classifiers for object detection include naive Bayes classifiers [10], Support Vector Machine (SVM) [11], neural networks [12], boosted decision stumps [13], *etc*. Here an optimal non-linear kernel-based SVM classification model will be investigated. SVM classifier proposed by [14] has recently received much attention in the machine learning community. It can be regarded as a linear model in a space defined by a non-linear mapping function Φ [15]: $f(x, \omega) = \omega^t \Phi(x) + b$, where ω corresponds to the model weights.

The trick of non-linear transformation $\Phi(x)$ is to project non-linear input spaces into more easily linearly separable high-dimensional feature spaces (Fig. 4). In soft margin method, the weights ω are obtained by resolving the optimization problem [16]:

$$\frac{1}{2}\|\omega\|^2 + \mathrm{C}\sum_{i=1}^{n}\xi_i \tag{2}$$

where ξ_i measures the degree of misclassification of the data x_i and parameter C determines the trade-off between the flatness of f and the error term [16]. Then the solution can be rewritten according to the kernel function K, defined by $K(x, x') = \langle\Phi(x), \Phi(x')\rangle$ as [15]:

$$f(x, \omega) = \sum_i \alpha_i y_i K(x_i, x) + b \tag{3}$$

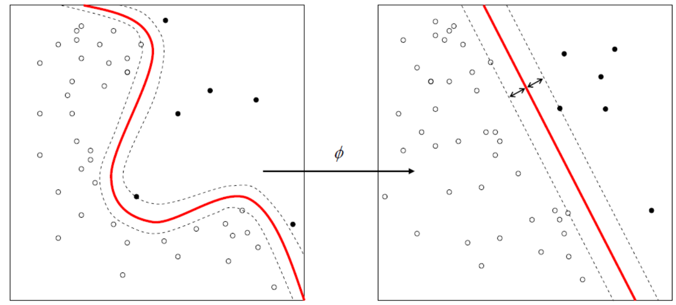

Fig. 4: Trick of non-linear transformation $\Phi(x)$ in SVM (Support Vector Machine)

where $\{(x_i, y_i)\}$ refers to the available training data, and α_i sets the relative weight of each training data in the classification model. The kernel function K can be seen as a measure of the degree of closeness of the test point x to the data base point x_i. Given a kernel model, the training of the SVR model resorts to the inference of the optimal weight vector according to a margin-based criterion. We let the reader refer to [15] for further details.

D. Deep learning

Instead of a two-stage architecture (feature extraction then classification step), we have intention to investigate a deep learning model. Deep learning is a new area of machine learning research and involves learned end-to-end architecture including trainable feature and trainable classifier [17]. It attempts to model high-level abstractions in data using multiple non-linear transformations. One of the promises of deep learning is replacing handcrafted features with efficient algorithms for unsupervised or semi-supervised feature learning and hierarchical feature extraction [18].

There are huge number of different variants of deep learning architectures. Convolutional neural networks are part of these architectures and are widely used for image and video recognition. It consists of multiple layers of small neuron collections which look at small portions of the input image, called receptive fields. The results of these collections are then tiled so that they overlap to obtain a better representation of the original image; this is repeated for every such layer. Convolutional networks may include local or global pooling layers, which combine the outputs of neuron clusters. We let the reader refer to [19], [20] for further details.

Caffe library is a fast open framework for deep learning [21]. It is developed by the Berkeley Vision and Learning Center (BVLC) and by community contributors. it is pure C++/CUDA architecture with Python and MATLAB bindings for training and deploying general purpose convolutional neural networks and other deep models architectures. It can be used to train a given labelled database to obtain classification models.

E. Object detection

Our objective of marine animals and birds detection is to determine if one or more instances of an object class are present and to find their locations and scales from input aerial image sequences. The most common approach to generic object detection is to slide a window across the image (at multiple scales if possible), and to classify each of them (local window) as containing the target or background with already trained classification model [22]. Another popular approach is to extract local interest points from the image, and then to classify each of the regions around these points, rather than looking at all possible sub-window [23]. The image segmentation techniques can help to find blob, classify image around detected blob at multiple scales.

IV. CONCLUSION

This paper proposed approaches to detect and classify targets on sea surface images registered by an airborne photography system. Two main groups of methods can be used in this domain. A first one is based on the extraction of features allowing to recognize each target using a classification model (Bayes, SVM, neural networks,...). A second approach consists in a deep learning system. The features and the mapping function are then included in the learning process. These methods allow to take into account the variety of targets (birds, marine mammals, boats, ocean phenomenons,...) and their different possible views. The choice of the classification model is then crucial and the efficiency of the classification will depend on it. These methods we have presented and developed in this paper will be then tested and compared on real data to validate our approaches.

ACKNOWLEDGMENT

This work is supported by ADEME, France, in collaboration with IN VIVO Environnement, LPO, the French Agence des Aires Marines Protégées, and Pixair. The authors thank Kelonia for providing images.

REFERENCES

[1] S. Buckland, M. Burt, E. Rexstad, M. Mellor, A. Williams, and R. Woodward, "Aerial surveys of seabirds : the advent of digital methods," *Journal of Applied Biology*, vol. 49, pp. 960–967, 2012.

[2] P. Terletzky and R. Ramsey, "A semi-automated single day image differencing technique to identify animals in aerial imagery," *Plos one*, vol. 9, January 2014.

[3] S. Gururatsakul, *Pattern recognition techniques for the identification of white sharks in aerial image sequences*. PhD thesis, 2011.

[4] I. Guyon and A. Elisseeff, "An introduction to feature extraction," in *Feature Extraction*, pp. 1–25, Springer, 2006.

[5] J. Canny, "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 6, pp. 679–698, 1986.

[6] T. Lindeberg, "Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention," *International Journal of Computer Vision*, vol. 11, no. 3, pp. 283–318, 1993.

[7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 886–893, IEEE, 2005.

[8] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2, pp. 1150–1157, Ieee, 1999.

[9] I. Levner, V. Bulitko, L. Li, G. Lee, and R. Greiner, "Automated feature extraction for object recognition," in *Proceedings of the Image and Vision Computing New Zealand conference*, pp. 309–313, 2003.

[10] S. Russell, P. Norvig, and A. Intelligence, "Artificial intelligence: A modern approach," *Prentice-Hall, Egnlewood Cliffs*, vol. 25, 1995.

[11] C. Papageorgiou and T. Poggio, "A trainable system for object detection," *International Journal of Computer Vision*, vol. 38, no. 1, pp. 15–33, 2000.

[12] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 1, pp. 23–38, 1998.

[13] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, vol. 4, pp. 34–47, 2001.

[14] V. Vapnik and A. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theory of Probability & Its Applications*, vol. 16, no. 2, pp. 264–280, 1971.

[15] B. Schölkopf and A. Smola, *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT press, 2001.

[16] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.

[17] Y. LeCun and M. Ranzato, "Deep learning tutorial," in *Tutorials in International Conference on Machine Learning (ICML13)*, Citeseer, 2013.

[18] Y. Bengio, "Learning deep architectures for ai," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[19] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[20] D. C. Ciresan, U. Meier, J. Masci, L. Maria Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," in *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, vol. 22, p. 1237, 2011.

[21] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[22] K. Murphy, A. Torralba, D. Eaton, and W. Freeman, "Object detection and localization using local and global features," in *Toward Category-Level Object Recognition*, pp. 382–400, Springer, 2006.

[23] R. Fergus, P. Perona, and A. Zisserman, "A sparse object category model for efficient learning and exhaustive recognition," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 380–387, IEEE, 2005.