

Comparaison de deux méthodes d'identification et de comptage automatique des espèces anguilliformes à partir de caméras acoustiques

AZENOR LE QUINIO¹, ALEXANDRE GIRARD², FABRICE ZAOUÏ¹, ERIC DE OLIVEIRA¹

¹ EDF R&D LNHE - Laboratoire National d'Hydraulique et Environnement

² EDF R&D PRISME - Performance, Risques Industriels et Surveillance pour la Maintenance et l'Exploitation
6 Quai Watier, 78401 Chatou, France

azenor.le-quinio@edf.fr, alexandre.girard@edf.fr, fabrice.zaoui@edf.fr, eric.de-oliveira@edf.fr

Résumé – Les caméras acoustiques sont fréquemment utilisées pour les suivis piscicoles du fait de leurs nombreux avantages pour l'observation des populations de poissons. Pour pallier l'analyse chronophage de leurs données effectuée par un opérateur, deux approches de traitements automatiques sont présentées basées sur différentes techniques de *computer vision*. La première repose sur de l'apprentissage supervisé et l'autre sur de l'analyse morphologique par traitements d'images. La comparaison de ces deux approches après mise au point et optimisation de chacune d'elles, montre une capacité de classification limitée pour la première (Score F1 = 54,2%) à l'inverse de la seconde qui permet d'obtenir de meilleurs résultats (Score F1 = 79,4%). Cette seconde approche arrive à pallier la qualité limitée des images traitées, notamment la fragmentation des images des espèces anguilliformes.

Abstract – Acoustic cameras are frequently used for fish monitoring because of their many advantages. To overcome the time-consuming analysis of their data carried out by an operator, two automatic processing approaches based on different computer vision techniques are presented. The first one is based on supervised learning and the other one on morphological analysis by image processing. The comparison of these two approaches, after development and optimization of each of them, shows a limited capacity of classification for the first one (F1-score = 54.2%) unlike the second one which makes it possible to obtain better results (F1-score = 79.4%). This second approach overcomes the limited quality of the processed images, especially the fragmentation of the images of anguilliform species

1 Introduction

La présence d'ouvrages le long des cours d'eau tend à faire obstacle aux mouvements migratoires indispensables aux cycles de vie de nombreuses espèces de poissons. Parmi elles, l'anguille européenne est une espèce catadrome (*c.-à-d.* vivant en rivière et se reproduisant en mer), aujourd'hui en voie critique d'extinction et qui fait donc l'objet de nombreuses politiques de conservations. L'étude de sa migration au niveau des ouvrages hydroélectriques est donc source de précieuses informations pour assurer sa préservation.

Différents outils sont aujourd'hui disponibles pour suivre ces populations de poissons. Parmi eux, on trouve les caméras acoustiques qui sont des sondeurs multifaisceaux. A l'inverse de sondeurs plus classiques qui produisent des échogrammes acoustiques, les caméras acoustiques enregistrent des flux d'images. Ces images sont reconstruites à partir des échos renvoyés par les objets se situant dans le faisceau de la caméra. Elles constituent une vue 2D de la section couverte par la caméra sur la largeur et la longueur du faisceau. Ces dispositifs ne permettent pas d'étudier les propriétés acoustiques des objets, mais permettent d'en observer la morphologie. Cependant, la principale limite à leur utilisation est l'analyse très chronophage de la grande quantité de données qu'elles collectent. Pour utiliser ces dispositifs au mieux, l'objectif est de développer une méthode d'identification et de comptage automatique des

espèces anguilliformes. Des études ont déjà été entreprises pour le comptage d'anguilles sur des données de caméras acoustiques, en utilisant des méthodes différentes. Des approches *deep learning* montrent des résultats encourageants sur de petits jeux de données [1], [2] mais plus limités sur des jeux de données avec une grande variabilité [3], ce qui est le cas des données traitées. Des méthodes de *computer vision* [4], [5] offrent également des résultats prometteurs, mais qui mettent en évidence le cas difficile et non pris en compte d'images d'anguilles fragmentées, quasiment systématiques dans les données de notre étude.

Le choix a donc été de développer deux approches indépendantes reposant sur des méthodes de classification différentes afin de voir celle qui répond au mieux au besoin. La première repose sur du *machine learning* tandis que la seconde s'appuie sur une analyse morphologique avec différentes méthodes de traitement d'images.

2 Traitements automatique proposés

2.1.1 Pré-traitement et suivi

Avant d'en effectuer le traitement, les fichiers de vidéos issues des caméras acoustiques des différents constructeurs sont préalablement convertis au format universel du conteneur Audio Video Interleaved. Plusieurs types de caméras acoustiques sont disponibles

dont les caractéristiques varient. Dans cette étude, nous utilisons une ARIS Explorer 1800 (Adaptive Resolution Imaging Sonar, Sound Metrics) installée pendant plusieurs mois sur la Dordogne, permettant ainsi d'utiliser directement des données *in situ* pour la mise au point de notre méthode.

Afin d'optimiser le temps d'analyse et d'effectuer un premier filtre de tous les objets dérivants passant dans le faisceau, une analyse globale de la vidéo est réalisée à l'aide d'échogrammes calculés à partir des valeurs singulières de chaque image. Pour chacune d'elles, les valeurs singulières sont calculées (1) sur des fenêtres glissantes et chevauchantes à 50%, et leurs ratios par rapport à une image de référence (2) en sont déduits. Pour éviter les effets de bord liés à la forme du faisceau de la caméra, un masque est calculé (3) et appliqué (4) à la matrice résultat de chaque valeur singulière. Puis le maximum de ces ratios de valeurs singulières est calculé selon la profondeur du champ de la caméra (5).

$$\forall w \in [1, W_1], \forall v \in [1, W_2],$$

$$\sigma_{w,v}^{N,*} = SVD^N(I_{i,j}^*)_{w * \frac{d_w}{2} \leq i \leq w * \frac{d_w}{2} + d_w, v * \frac{d_w}{2} \leq j \leq v * \frac{d_w}{2} + d_w} \quad (1)$$

$$S_{w,v}^N = \frac{\sigma_{w,v}^N}{\sum_{n=1}^N \sigma_{w,v}^N} - \frac{\sigma_{w,v}^{N,ref}}{\sum_{n=1}^N \sigma_{w,v}^{N,ref}} \quad (2)$$

$$M_{w,v} = \min(I_{i,j}^{ref})_{w * \frac{d_w}{2} \leq i \leq w * \frac{d_w}{2} + d_w, v * \frac{d_w}{2} \leq j \leq v * \frac{d_w}{2} + d_w} \quad (3)$$

$$R_{w,v}^N = S_{w,v}^N * M_{w,v} \quad (4)$$

$$\forall v \in [1, W_2], V_{v,1}^N = \max(R_{w,j}^N)_{1 \leq w \leq W_1} \quad (5)$$

Avec W_1 et W_2 le nombre de fenêtres glissantes suivant les axes horizontaux et verticaux ; I , l'image étudiée ; I^{ref} , l'image de référence ; d_w , la dimension de la fenêtre d'étude ; $\sigma_{w,v}^N$, la $N^{\text{ième}}$ valeur singulière de la fenêtre (w, v) ; $\sigma_{w,v}^{N,ref}$, la $N^{\text{ième}}$ valeur singulière de la fenêtre (w, v) de référence ; S , la matrice des ratios ; M , le masque ; R , la matrice résultat des ratios.

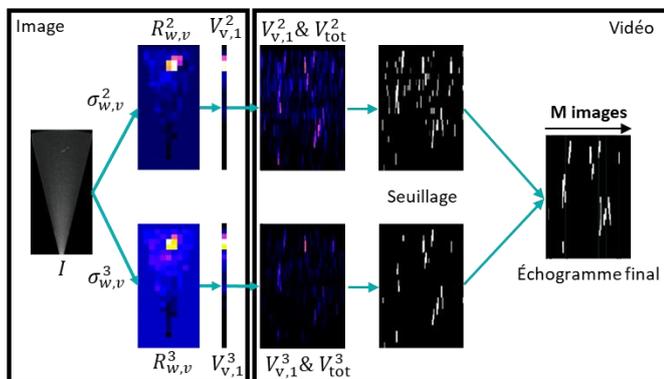


Figure 1 : Etapes du calcul de l'échogramme de discontinuités

Ces échogrammes permettent d'étudier les discontinuités présentes sur l'ensemble de la profondeur de la caméra, perpendiculairement à l'axe du courant, tout en compressant les informations pour plus d'efficacité. Cette méthode de calcul est une alternative à

celle utilisée par le logiciel ARIS Fish software qui utilise la valeur maximale de l'intensité des échos [6]. Les échogrammes des 2^{ème} et 3^{ème} valeurs singulières sont ensuite fusionnés et les fortes discontinuités communes isolées pour récupérer les intervalles d'images correspondants (Figure 1).

Les images extraites sont ensuite analysées afin de permettre la classification des objets qu'elles présentent. Un filtre moyenneur est d'abord réalisé, suivi d'une soustraction d'arrière-plan reposant sur un mélange gaussien d'après [7]. Afin d'identifier les régions à analyser et de mieux prendre en compte le cas des images d'anguilles fragmentées, une dilatation [8] de l'image est appliquée. Chaque objet ainsi isolé est extrait dans une plus petite image de taille fixe à partir de laquelle seront réalisées les analyses selon la méthode choisie.

Afin de réaliser un comptage correct des objets classifiés, leur suivi tout au long de leur passage dans le faisceau de la caméra est systématiquement mené. Le voisinage de l'objet est étudié durant un court laps de temps ($\Delta t = 0.7s$) suivant sa dernière détection. Ainsi, si un objet est détecté dans la zone de recherche, avec une orientation similaire à la détection précédente, il sera considéré comme la suite de ce suivi et cette zone sera mise à jour et centrée autour de cette nouvelle détection. A l'inverse, si un objet est détecté, mais ne se situe dans aucune zone de recherche, un nouveau suivi indépendant sera lancé. Plusieurs suivis peuvent être ainsi réalisés en même temps.

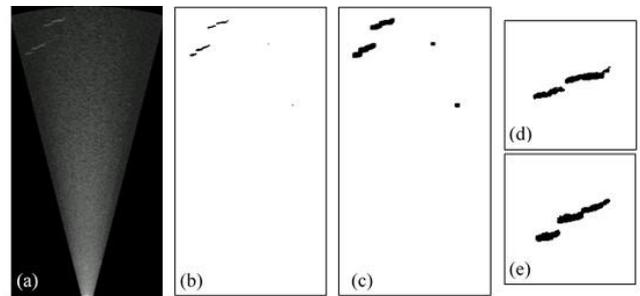


Figure 2 : a) Image traitée par filtre moyenneur ; b) Image binaire obtenue par soustraction d'arrière-plan ; c) Dilatation de l'image binaire ; d,e) Extrait des deux candidats

2.1.2 Classification par machine learning

Le modèle d'apprentissage utilisé repose sur une structure du type descripteur-classifieur. Il a nécessité la création d'une base de données qui contient des images d'anguilles européennes et des images de débris et d'autres espèces de poissons non anguilliformes. Ces images sont de taille fixe, et centrées autour de l'objet qu'elles présentent. Au total, près de 45.000 images ont été labellisées entre les deux classes (« Anguilliformes », « Autres ») avec respectivement 46,5% pour la première classe et 53,5% pour la seconde. Des méthodes de *data augmentation* par rotation et inversion ont été utilisées afin d'obtenir ce nombre équilibré d'images.

Le descripteur choisi pour concevoir cette approche est l'histogramme de gradient orienté. Il permet de prendre

en compte la magnitude du gradient, mais aussi les variations locales d'orientation.

Pour ce qui concerne le classifieur, trois ont été testés, à savoir la méthode des machines à vecteurs de supports (SVM), la méthode des K-voisins (KNN) et celle de la *Random Forest* (RF). Avant d'en évaluer la performance sur les données de notre étude, chacun d'eux a été optimisé en utilisant la méthode de validation croisée K-folds. Les paramètres analysés dépendent du classifieur choisi. Si l'on prend le cas du SVM, le couple (C, γ) est étudié, correspondant aux paramètres du noyau Gaussien radial du SVM choisi.

À la vue des performances obtenues par chacun des classifieurs testés (Tab 1.), le plus adéquat pour notre étude est le SVM. Il permet d'obtenir des taux d'exactitude, de précision et de rappel de plus de 90%.

Chaque image de l'objet détecté et étudié par la suite sera confrontée à ce modèle afin qu'il détermine si l'objet est considéré comme une espèce anguilliforme ou non. Sur l'ensemble des détections d'un même objet, celui-ci doit avoir été classifié comme anguilliforme à au moins cinq reprises pour être finalement identifié comme tel.

Tab 1 : Performance théorique des trois classifieurs

	Exactitude	Précision	Rappel
SVM	93,1%	91,2%	94,4%
RF	87,9%	84,6%	90,3%
KNN	89,9%	86,5%	93,2%

2.1.3 Classification par analyse morphologique

Cette méthode-ci utilise une succession de techniques de traitement d'images afin de réaliser une analyse de la morphologie des objets à identifier. Plusieurs caractéristiques physiques sont ainsi extraites à chaque détection de l'objet : la surface, l'excentricité globale, la longueur ainsi qu'une caractérisation de la silhouette.

Le calcul de la taille des individus est réalisé à partir du squelette de leurs corps [10]. Du fait du fractionnement fréquent de l'image d'un même objet, une reconstruction rapide de ce dernier est menée afin de récupérer une information la plus fiable possible. La silhouette uniformément fine, très caractéristique des espèces anguilliformes, est également étudiée comme l'un des éléments permettant la classification. L'image binaire de l'objet est segmentée en trois parties (Figure 3) par la méthode des k-moyennes [9], avec $k = 3$. L'excentricité de chacune d'elle est ensuite mesurée. Ce processus permet d'analyser l'allongement de l'objet et son homogénéité, qu'importe la courbure du corps.



Figure 3 : Exemples de fragmentation et de calculs d'excentricité pour une espèce anguilliforme (a) et non anguilliforme (b)

Chaque détection d'un même objet est soumise à ce même processus permettant ainsi de prendre en compte

l'évolution de ses propriétés morphologiques au cours de son mouvement. Des informations sur la dynamique de l'objet (vitesse et forme du déplacement) sont aussi récupérées grâce au suivi complet de son mouvement dans le faisceau de la caméra. Elles permettent ainsi d'ajouter des informations complémentaires à la classification.

L'identification de l'objet en espèce anguilliforme repose sur un ensemble de seuils empiriquement établis prenant en compte l'ensemble des informations extraites. Ainsi, sont considérés comme anguilliformes, les objets répondant positivement à l'ensemble de ces critères. Les seuils de classification utilisés dépendent de certaines propriétés d'enregistrement de la caméra tels que le nombre de frames par seconde et la résolution des pixels.

3 Performances obtenues

Les performances des deux approches mises au point ont été estimées à partir des comptages visuels d'un ensemble de vidéos, de passages d'anguilles européennes, qui a servi de référence. Ces comptages consistent à visionner chaque vidéo par des observateurs afin de récupérer l'heure de chaque passage. Bien qu'il existe toujours un biais opérateur, le nombre de passages d'anguilles par vidéo étant faible, ce biais reste, dans le cadre de cette étude, limité [11].

Les résultats des deux approches obtenus sur les mêmes données ($N = 170$ passages d'anguilles) montrent des résultats variant conséquemment d'une approche à l'autre. Les taux de rappel (6) obtenus sont de 53,5% et de 77,1% pour l'approche *machine learning* et l'approche d'analyse morphologique respectivement. Similairement, leurs taux de précision (7) atteignent 54,8% et 81,9%.

$$Rappel = \frac{VP}{VP+FN} \quad (6) \quad Précision = \frac{VP}{VP+FP} \quad (7)$$

Avec VP, le nombre de vrais positifs ; FN, le nombre de faux négatifs ; FP, le nombre de faux positifs.

Dans les deux cas, les régressions linéaires (Figure 4) entre les comptages automatiques et ceux de référence suivent une pente inférieure à 1 avec cependant une forte corrélation observée pour la seconde approche ($R^2 \geq 0.91$).

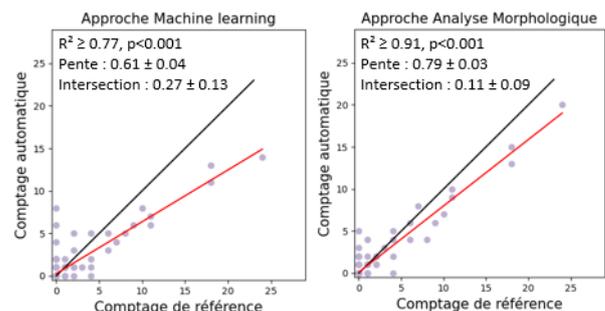


Figure 4 : Graphique des régressions linéaires entre comptage de référence et comptage automatique pour chaque approche

Afin de compléter l'analyse de cette deuxième méthode qui semble davantage prometteuse, nous l'avons testée à plus large échelle, sur 753 passages d'anguilles. Un rappel de 73,8% et une précision de 83,5% sont ainsi obtenus. Parmi les principales causes d'erreurs, on observe la confusion avec d'autres espèces de poissons (37,3%) et la confusion lors de phénomènes de « arcing » (36,4%), qui sont des déformations observées lors de la reconstruction de l'image à partir des échos. Ils étirent de façon longiligne l'image de certains objets [5], tout particulièrement ceux proches de la caméra et de fortes intensités.

4 Conclusion

Notre étude a ainsi cherché à mettre au point une méthode d'identification et de comptage des espèces anguilliformes à partir des données issues de caméras acoustiques. Deux approches ont été développées, basées sur des techniques de classification différentes. La première repose sur du *machine learning* et la seconde sur de l'analyse morphologique par traitement d'image. Avec un taux de rappel et une précision qui ne dépassent pas tous deux les 60%, la première approche ne permet pas de fournir des comptages suffisamment fiables pour leurs exploitations dans le cadre de suivis piscicoles. A l'inverse, la seconde approche permet de comptabiliser près des trois-quarts des anguilles passant dans le faisceau de la caméra tout en limitant les fausses détections à 2 comptages sur 10. Bien qu'elle tende à sous-estimer le nombre total de passages d'anguilles, cette approche permet d'observer les dynamiques de migration des espèces étudiées, notamment avec l'identification des grands pics de dévalaison et de montaison, qui sont des informations précieuses pour le suivi de ces espèces. La principale difficulté rencontrée par la classification par *machine learning* est liée à la fragmentation fréquente des images de poissons, tout particulièrement chez les espèces anguilliformes. Leur long corps marqué par une forte ondulation rend fréquente une représentation discontinue [1], [5] de leur image par la caméra acoustique. En effectuant une analyse morphologique plus poussée, dans la seconde approche, nous réussissons à pallier davantage cette limite, bien qu'elle puisse également être source d'erreurs notamment à cause des phénomènes d'arcing dont un pré-traitement spécifique mériterait d'être mis en place afin de limiter les erreurs engendrées. De plus, pour pallier la confusion avec d'autres espèces, il serait intéressant d'étudier les caractéristiques morphologiques de ces espèces afin d'améliorer les seuils de décision.

L'analyse morphologique proposée nous semble donc être l'approche la plus adaptée à notre besoin d'automatisation du traitement des données issues de caméras acoustiques. Elle présente en plus d'une bonne performance sur la caméra ARIS, des résultats similairement encourageants (Rappel = 70,2%, Précision = 74,3% pour N = 198 passages) sur un second type de

caméra, la BlueView 2D (BlueView Technologies, Inc. SubSeaTech s.r.l.). Cette dernière est plus récente et tend à être davantage utilisée du fait de son ouverture de champ à l'horizontale quatre fois plus large. Ces travaux méritent aujourd'hui d'être généralisés à d'autres types d'espèces afin de permettre une application plus large.

Références

- [1] T. Yin, X. Zang, Z. Hou, P. Jacobson, R. Mueller, and Z. Deng, 'Bridging the Gap between Laboratory and Field Experiments in American Eel Detection Using Transfer Learning and Convolutional Neural Network', 2020. doi: 10.24251/HICSS.2020.116.
- [2] X. Zang, T. Yin, Z. Hou, R. P. Mueller, Z. D. Deng, and P. T. Jacobson, 'Deep Learning for Automated Detection and Identification of Migrating American Eel *Anguilla rostrata* from Imaging Sonar Data', *Remote Sensing*, vol. 13, no. 14, 2021, doi: 10.3390/rs13142671.
- [3] G. Fernandez Garcia, F. Martignac, M. Nevoux, L. Beaulaton, and T. Corpetti, 'A deep neural network for multi-species fish detection using multiple acoustic cameras', Sep. 2021. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-03350565>
- [4] L. Bothmann, M. Windmann, and G. Kauermann, 'Realtime classification of fish in underwater sonar videos', *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 65, p. n/a-n/a, 2016, doi: 10.1111/rssc.12139.
- [5] A.-M. Mueller, T. Mulligan, and P. Withler, 'Classifying Sonar Images: Can a Computer-Driven Process Identify Eels?', *North American Journal of Fisheries Management - NORTH AM J FISH MANAGE*, vol. 28, pp. 1876–1886, 2008, doi: 10.1577/M08-033.1.
- [6] Sound Metrics Corp, 'ARISFish Software User Guide: Echogram Counting Tutorial'. Sound Metrics Corp, 2014.
- [7] Z. Zivkovic and F. Van der Heijden, 'Efficient adaptive density estimation per image pixel for the task of background subtraction', *Pattern Recognition Letters*, vol. 27, pp. 773–780, 2006, doi: 10.1016/j.patrec.2005.11.005.
- [8] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. USA: Prentice-Hall, Inc., 2006.
- [9] S. P. Lloyd, 'Least squares quantization in PCM', *IEEE Trans. Inf. Theory*, vol. 28, pp. 129–136, 1982.
- [10] H. Lü and P. Wang, 'A comment on a fast parallel algorithm for thinning digital patterns', *Communications of the ACM*, vol. 29, pp. 239–242, 1986, doi: 10.1145/5666.5670.
- [11] R. Lagarde *et al.*, 'In situ evaluation of European eel counts and length estimates accuracy from an acoustic camera (ARIS)', *Knowledge and Management of Aquatic Ecosystems*, vol. 421, p. 44, 2020, doi: 10.1051/kmae/2020037.