Detect, track and identify fish : image processing to automate acoustic camera data analysis in shallow waters

Azénor Le Quinio

COLLEGE ECOLOGIE
DOCTORAL GEOSCIENCES
BRETAGNE AGRONOMIE ALIMENTATION

L'INSTITUT agro Rennes Angers

THESE DE DOCTORAT DE

L'INSTITUT AGRO RENNES ANGERS

ECOLE DOCTORALE N° 600
Ecologie, Géosciences, Agronomie, Alimentation
Spécialité : *Statistiques/ Modélisation en écologie, géosciences, agronomie et alimentation*

Par
Azénor LE QUINIO

Detect, track and identify fish: image processing to automate acoustic camera data analysis in shallow waters

Thèse présentée et soutenue à Chatou, le 28/06/2023
Unité de recherche : EDF R&D (LNHE) / INRAE (UMR DECOD)
Thèse N° : D-102- 2023-12

Rapporteurs avant soutenance :

Olle Calles	Professeur, Karlstad University, Suède
Thomas Walter	Professeur, Directeur du CBIO, Mines ParisTech, France

Composition du Jury :

Président :	Kacem Chehdi	Professeur, ENSSAT Lannion, Université de Rennes 1, France
Examinateurs :	Ana Da Silva	Directrice de recherche, Norwegian Institute for Nature Research, Norvège
	Mathieu Emily	Professeur, L'institut Agro Rennes-Angers, France
	Elsa Goerig	Associée de recherche, Harvard Museum of Comparative Zoology, Etats-Unis
Dir. de thèse :	Jean-Marc Roussel	Directeur de recherche, UMR DECOD, INRAE, France
Encadrement de thèse :	François Martignac	Ingénieur de recherche, UMR DECOD, INRAE, France

Invité(s)

Eric De Oliveira	Ingénieur de recherche, EDF R&D, France
Jean Guillard	Ingénieur de recherche, UMR CARRTEL (INRAE - USMB), France

Titre : Détection, suivi et identification des poissons : le traitement d'images pour automatiser l'analyse des données des caméras acoustiques en eaux peu profondes

Mots clés : Imagerie acoustique, Traitement d'images, Migration piscicole, Suivi de population

Résumé : Les caméras acoustiques sont des outils de plus en plus utiliser pour les suivis piscicoles. Non-intrusives, elles permettent d'effectuer des suivis continus, sans aucune manipulation des poissons, avec une portée supérieure aux appareils optiques grâce à leur faible dépendance aux conditions environnementales. Le développement d'une méthode automatique afin de détecter, compter et identifier les poissons à partir des images acoustiques est un enjeu clé pour exploiter pleinement les capacités des caméras acoustiques. De nombreuses caractéristiques rendent possible l'identification des espèces de poissons, mais nombre d'entre elles ne peuvent être extraites des caméras acoustiques en raison de la nature et de la résolution des vidéos qu'elles enregistrent. Cela fait de l'analyse automatique des vidéos acoustiques un véritable défi. Ce travail de thèse vise ainsi à développer une méthode automatique d'identification des espèces de poissons ou des groupes d'espèces grâce à l'extraction des caractéristiques de morphologie, de mouvement et de mode de nage, en se basant sur des outils de traitements d'images.
Ce travail permet d'extraire l'ensemble des informations disponibles à partir des vidéos acoustiques, constituant une avancée importante pour le développement d'une méthode automatique d'identification des espèces. Ce travail a également démontré que l'intégration des conditions environnementales et des connaissances écologiques est nécessaire pour mener à bien ce processus. La mise au point d'un modèle de classification intégrant l'ensemble de ces informations est la prochaine étape à réaliser.

Title: Detect, track and identify fish: image processing to automate acoustic camera data analysis in shallow waters

Keywords: Acoustic imaging, Image processing, Fish migration, Population monitoring

Abstract: Acoustic cameras are increasingly used in fish population monitoring studies. Non-intrusive, they allow continuous fish monitoring without any manipulation of the fish while covering a range superior to optical devices due to their low dependence on environmental conditions. The development of an automatic method to detect, count and identify fish from acoustic images is a key issue to fully exploit the capabilities of acoustic cameras. Many features can be used to identify fish species, but many cannot be extracted from acoustic cameras due to the nature and resolution of the video they record. This makes the automatic analysis of acoustic videos a real challenge. The aim of this PhD work is to develop an automatic method for identifying fish species or groups of species through the extraction of morphology, motion and swimming characteristics, based on image processing tools. This work allows the extraction of all the information available from the acoustic videos, which constitutes an important step for the development of automatic species identification methods. This work has also shown that the integration of environmental conditions and ecological knowledge is necessary to carry out this process. The development of a classification model integrating all of this information is the next step to be taken.

Acknowledgements

Cette thèse n'aurait pas été la même et n'aurait sans nul doute pas aboutie sans les personnes bienveillantes qui m'ont entourées durant ces trois années. Je tiens donc à vous remercier avec ces quelques mots.

Avant toute chose, je souhaite remercier l'ensemble des membres de mon jury d'avoir accepter d'évaluer mes travaux de thèse à travers leurs regards de chercheur(e)s averti(e)s. Cela a été un réel plaisir et une chance de pouvoir vous présenter mes travaux et d'échanger avec vous au cours de ma soutenance. Merci donc aux Pr. Kacem Chehdi, Pr. Mathieu Emily, Dr. Ana Da Silva et Dr. Elsa Goerig. Merci également à mes deux rapporteurs Pr. Olle Calles et Pr. Thomas Walter.

Je souhaite ensuite remercier mon encadrement de thèse. J'ai bien conscience d'avoir eu beaucoup de chance d'avoir une équipe encadrante comme celle que j'ai eu et je tiens sincèrement à vous remercier pour votre temps, votre bienveillance et votre patience.

Jean-Marc, je suis très reconnaissante de t'avoir eu pour directeur de thèse. Ta vision globale sur mes travaux a permis d'avoir le recul nécessaire à leur bon avancement. Ton expérience et ta rigueur scientifique ont également beaucoup apporté à ma formation de future chercheure.

Jean, je souhaite te remercier d'avoir co-dirigé ma thèse. Tes remarques, toujours pertinentes, m'ont encouragées à être toujours plus rigoureuse mais aussi à questionner les différents chemins envisagés pour mes travaux.

François, je tenais à te remercier pour ces trois années passées sous ton encadrement. Ta bienveillance, ta volonté à mettre en avant mes travaux et ta positivité ont été d'un grand réconfort et d'une grande motivation. Bien que mes passages à Rennes n'aient pas été aussi nombreux que prévu, ta disponibilité à mon égard lors de chacun d'entre eux m'a beaucoup apporté tant sur le plan scientifique que sur le plan humain, tout comme ton investissement au quotidien, même à distance. Je tiens également sincèrement à te remercier pour toute l'énergie que tu as consacré à mon premier article, exercice qui n'a pas été simple pour moi et que ton investissement a permis de rendre plus facile tout en m'apprenant énormément.

Eric, un immense merci pour ces trois ans de thèse mais également pour ces quasiment quatre ans de travail ensemble. Elle me semble bien loin cette première discussion où tu m'as présenté le sujet qui allait devenir, mon, sujet de thèse. Depuis il y a eu beaucoup de chemin parcouru et si cette thèse a aboutie tu n'y es pas pour rien. Nos discussions m'ont beaucoup apportées d'un point de vue scientifique et m'ont souvent permis de trouver de nouvelles idées à tester. Mais surtout, merci pour ton soutien au quotidien, de m'avoir remotivé dans les moments compliqués de la thèse, d'avoir écouté mes doutes et mes désaccords et d'avoir cherché à les résoudre pour que j'aille au bout.

Je tiens à remercier également les personnes qui, bien que ne faisant pas officiellement partie de mon encadrement de thèse, y ont contribué grandement. Alexandre, merci pour le temps que tu as consacré au quotidien à ma thèse. Tes importantes connaissances scientifiques ainsi que nos nombreux échanges m'ont beaucoup aidé pour réaliser mes travaux. Je te suis très reconnaissante d'avoir suivi ma thèse et de m'avoir conseillé tout au long de ces trois années. Fabrice, cela a été un plaisir de travailler et d'échanger avec toi. Ton expertise et ta volonté de partager tes connaissances m'ont beaucoup apporté autant scientifiquement qu'humainement. Je tenais donc à te remercier pour le temps que tu as consacré à nos discussions et pour l'aide que tu m'as donnée.

Je tiens également à remercier les personnes avec qui j'ai pu échanger pendant ma thèse et notamment l'équipe du CEFREM de l'Université de Perpignan qui ont notamment accepté de nous mettre à disposition une partie de leurs données. Merci également à Younès pour les échanges que nous avons eu, avoir participé à ton encadrement de stage a été l'une des belles expériences de ma thèse.

Durant ces trois années j'ai été hébergée à l'INRAE comme à EDF et je tenais donc à remercier les équipes qui m'ont accueillie. Merci aux membres de l'équipe de l'UMR DECOD de Rennes pour leur accueil et leur bienveillance pendant mes séjours chez eux.

Merci également à l'équipe P76 du département LNHE mais aussi à l'ensemble des agents du bâtiment Q. Cela a été un plaisir de faire ma thèse parmi vous. L'environnement dans lequel se déroule une thèse est un facteur important de son bon déroulement alors vous y êtes pour beaucoup ! Merci à Sylvie dont le projet Biodiv a hébergé ma thèse ainsi qu'aux deux chefs de groupes EDF qui m'ont accueillies dans leur équipe, Sylvain et Jérémy. Merci à Rachid pour tous ces disques durs ! Une petite dédicace à ceux qui m'auront fait beaucoup rire et mise de bonne humeur qu'importe les moments traversés, Mathieu et Christine. A celles dont la bienveillance, les conseils, et le soutien m'auront beaucoup apporté pendant ma thèse, Julie, Sabine et Laurence. A l'équipe EMMN avec qui je suis partie à Poses et avec qui j'ai partagé

les délicieux déjeuners de la guinguette, Samy, Matthieu et Sébastien. Merci à toutes et tous pour les échanges à la salle café (qui a été mon deuxième bureau en attendant ma soutenance), Thibaud, Antonin, et bien d'autres. Merci à celles et ceux avec qui j'ai partagé de très bons moments (très souvent autour d'un verre) depuis mon arrivée au LNHE, Laure, Anthony (on se fait une revanche au baby-foot quand tu veux !), Léo (tu seras toujours mon binôme de sorties post-boulot !), Nastassia et Florian. J'ai été très heureuse de faire ces trois ans à vos côtés ! Merci également à Romain pour toutes nos discussions, cela a été tout autant un plaisir d'échanger sur mes travaux de thèse avec toi que motivant d'en voir leur utilité.

Merci également à la team des « non-statutaires » du LNHE, Mathilde, Erika, Coline, Andrès, Sarah, Noor, Sophie, Théo, Flora, Helloïse et Charles. Un grand merci à Mathilde, ma co-bureau sur le finish. Merci de m'avoir supportée lors de ma période de rédaction, ta gentillesse et ta bonne humeur ont été une réelle bouffée d'air frais au quotidien. L'ambiance de notre bureau va sincèrement me manquer. Merci également à mon premier co-bureau, Nicolas. Ta jovialité, ton humour et ton expérience de thèse ont fait du bureau un environnement propice à mon début de thèse et y sont pour beaucoup dans mon intégration au LNHE. Merci à toi Erika pour ton soutien et ta capacité de motivation. Merci d'avoir été toujours partante pour aller décompresser après le travail, étape non négligeable de la journée ! Je ne désespère pas qu'on aille un jour se faire un petit cours de zumba sur la plage ! Merci Andrès pour ta bonne humeur communicative et ta gentillesse. Nous voilà enfin Docteur et Docteure et je suis heureuse que tu sois de retour à Paris pour qu'on partage nos nouvelles péripéties de chercheur(e) ! Merci Coline pour ces trois années de thèse passées ensemble. Ce fût un réel réconfort de partager toutes les étapes de nos thèses et d'avoir pu s'entraider tout au long. Je t'avoue que j'ai du mal à réaliser que cette aventure se termine ! Merci également Sarah pour ta gentillesse et ta bienveillance, t'avoir en voisine de bureau a été un vrai plaisir. Tu as ouvert la marche des soutenances en nous donnant un très bel exemple ! Merci Sophie pour ton énergie et ta bonne humeur débordante. La porte de mon bureau te sera toujours grande ouverte, profite à fond de tes trois ans de thèse, je crois que je peux aujourd'hui te dire que c'est une très belle aventure !

Je ne peux citer tout le monde mais le cœur y est sincèrement !

Je tiens également à remercier les doctorants et agents du bâtiment S du département PRISME qui m'ont accueillie lors de mon stage de fin d'étude. Avant ce stage je n'envisageais pas du tout de faire une thèse, je l'ai néanmoins fini avec un sujet dans les bras. Votre enthousiasme, votre motivation et votre volonté de partage vis-à-vis de vos thèses et autres sujets de recherche m'auront convaincus de me lancer à mon tour et pour cela je vous en suis très reconnaissante. Nos échanges ces trois dernières années m'auront été tout autant enrichissants. Merci donc à Paul, Thomas, Jérôme, Alvaro, Laura, Julien et Vincent. Merci également à mes deux tuteurs de stage, Claire et Nicolas, c'est grâce à vous que je suis arrivée à la R&D d'EDF et

grâce à vous également que j'ai entendu parler de mon sujet de thèse. Merci de m'avoir partagé votre expérience et vos connaissances scientifiques. Merci à Guy de m'avoir parler de ce sujet de thèse et mis en contact avec Eric, c'est grâce à toi que j'écris aujourd'hui les remerciements de mes travaux.

Je voudrais adresser également un merci tout particulier aux stagiaires de la promo 2019 avec qui j'ai partagé également ces trois dernières années de thèse, Mathilde, Pierre, Antoine, Vanessa et Clément ! Vanessa, merci pour ton éternel soutien et ta gentillesse. Tu es depuis le début une oreille attentive et bienveillante sur laquelle j'ai pu me reposer et dont les conseils sont précieux. Tu fais partie de ces personnes qui en peu de mots mettent de bonne humeur et je t'avoue que tu nous manque à tous depuis ton départ à Toulouse (mais je n'écarte pas l'idée d'y migrer à mon tour un jour) ! Antoine, merci pour ta bonne humeur, ton calme et ta positivité qui permettent de remettre les choses en perspective. Tes anecdotes m'auront beaucoup faire rire ces trois dernières années. Je tiens également à dire que tu es un très bon binôme de pétanque (même si l'expérience nous a montré que pour gagner il faut toujours une bretonne) !

Merci également aux doctorants de cette promotion Pablo, Théo, Marouane et Elias. Merci pour votre soutien, votre humour et votre bonne humeur communicative qui ont rendu ces trois ans de thèse tellement plus agréable. Un merci tout particulier à toi Sami. Merci pour ton soutien, ta bonne humeur, tes blagues et tes conseils qui m'ont été précieux. Je te serais toujours très reconnaissante d'avoir été présent les derniers mois de ma thèse, on ne va pas se mentir beaucoup de choses se sont accumulées alors merci d'avoir fait en sorte que la cocotte-minute n'explose pas !

Dans une thèse on embarque donc, un peu malgré eux, nos proches et je tenais à les remercier sincèrement. Merci à celles et ceux qui ont, à un moment ou un autre de ces trois années, partagé avec moi un bout de ma thèse : de la publication de mon premier article, aux retours de mes rapporteurs et jusqu'à mon titre de Docteure ! Un merci tout particulier à toi Céline pour ton amitié et ton éternel soutien depuis maintenant quatorze ans (ça ne nous rajeunit pas tout ça) ! Et à toi, Ismaël pour ton soutien et ta sollicitude durant ces trois ans, ça en fait du chemin depuis Hong Kong !

Merci à mes parents, Thierry et Corinne, pour leur soutien, tout particulièrement cette dernière année. Merci d'avoir été là lorsque j'avais besoin de décompresser et de conseils. Sachez que soutenir ma thèse devant vous était un moment très important pour moi. Merci également à ma sœur et mon frère, Maïwenn et Youwann. Merci Maïwenn d'avoir été la première à me dire de foncer et que si le sujet me plaisait, je n'avais qu'à la faire cette thèse ! Tu as également droit à une dédicace toute particulière pour toutes tes relectures d'anglais. Je crois

que l'on peut dire aujourd'hui que les anguilles n'ont plus de secret pour toi ! Merci Youwann, te voir t'intéresser à mon sujet et mes travaux valait la reconnaissance de bien du monde. De vous quatre tu es celui qui est le plus sensibilisé à ce monde de la recherche et si un jour tu te lances dans une thèse, sache que tu pourras compter sur ma longue expérience de doctorante et que je suis prête à partager le titre de Dr. Le Quinio ! Merci également à celles et ceux qui m'ont soutenus pour ma soutenance, Emmanuel, Rozen, Karine et Stéphane. Cette thèse est aussi l'aboutissement de huit années d'études supérieures et je tiens à remercier une personne à qui je dois beaucoup et à qui je n'aurais malheureusement pas l'occasion de le dire de vive voix, je le fais ici, merci Papi.

J'espère que la lecture de ces remerciements auront fait sourire celles et ceux qui se reconnaîtront. Je tiens humblement à faire remarquer qu'ils me semblent plus réussis, bien que tout aussi long, que les remerciements de ma soutenance !

Ces trois années de thèse s'achèvent donc. Elles auront été riches en émotions et en apprentissage, et je dois dire que c'est avec soulagement et avec joie que je laisse ce manuscrit entre les mains de ses lecteurs et que je clôture ce beau chapitre.

Bonne lecture,

Azénor

Contents

List of Figures

List of Tables

List of Abbreviations

AC Acoustic Cameras.

ARIS Adaptive Resolution Imaging Sonar.

ARIS 1800 ARIS Explorer 1800.

ARIS 3000 ARIS Explorer 3000.

BCF Body and/or Caudal Fin.

BRIEF Binary Robust Independent Elementary Features.

BS Background Subtraction.

BV BlueView M900-2250-130 2D.

DIDSON Dual-frequency Identification Sonar.

DoG Difference of Gaussian.

DTW Dynamic Time Warping.

EDF Électricité de France.

eDNA environmental DNA.

FAST Features from Accelerated Segment Test.

FEM Finite-Elements Methods.

FN False Negatives.

FOI Frames Of Interest.

FOV Field Of View.

FP False Positives.

FPS Frames Per Second.

HF High-Frequency.

HoG Histogram of Oriented Gradients.

KDE Kernel Density Estimation.

LF Low-Frequency.

MHT Multiple Hypothesis Tracking.

MoG Mixture of Gaussians.

MPF Median and/or Paired Fin.

MSM Mass-Spring Model.

OOI Objects Of Interest.

ORB Oriented FAST and Rotated BRIEF.

PDF Probability Density Function.

ROI Regions Of Interest.

SIFT Scale-Invariant Feature Transform.

SURF Speeded Up Robust Features.

SV Singular Value.

SVD Singular Value Decomposition.

SVR Singular Value Ratio.

TP True Positives.

Introduction

Contents

La théorie, c'est quand on sait tout
et que rien ne fonctionne.
La pratique, c'est quand tout fonctionne
et que personne ne sait pourquoi.
Albert Einstein

1 Migration and ecological continuum: a challenge for fish population preservation

1.1 Migration: a vital movement to complete fish life cycle

Migration is a persistent and straightened-out movement achieved by the animal's own locomotor means or by actively seeking a transport vector, such as air or water currents, during which it remains undistracted by the resources it might find during migration by temporarily inhibiting 'station-keeping responses', and which might be repeated later in life [Morais and Daverat, 2016], as defined by [Kennedy, 1985] and discussed by [Dingle, 1996]. Migration (Figure 1) varies both between and within species [Chapman et al., 2012] and therefore takes many different forms depending on the organism, the spatial or the temporal attributes as well as the medium it uses [Dingle and Drake, 2007]. The causes of migration are diverse but the most common one remains the search for resources, especially food [Dingle and Drake, 2007]. Hence, migration is usually considered as an adaptive response to the seasonal and/or geographic variation of food abundance or quality [Gauthreaux, 1982]. However, other causes such as spawning, seasonal refuge from predators or adverse environmental conditions, might also lead to migration [Brönmark et al., 2013]. Overall, a migration behaviour of an organism is led by the objective of completing its life cycle.

Regarding freshwater fish population, species display different spawning, feeding and growing habitats preferences that support different migration strategies. The decision to migrate or not is made by the individual and is part of an adaptive strategy to maximise lifetime reproductive effort [Brönmark et al., 2013]. Migration strategies can be sorted depending on traveling distances and habitats explored. Many freshwater fish migrate within the watershed including main river stem or lake and tributaries, they are called potadromous species. As an example, the Northern Pike (*Esox Lucius*) performs spawning migration from the lakes to their tributaries or from main rivers to floodplain hydro systems, being able to travel tens of kilometers [Keith et al., 2021]. Other species complete their migration between freshwater and marine environments, they are called the diadromous species. Many of them reproduce in freshwaters, before juveniles join the marine environment to grow and start sexual maturation. They finally perform their upstream migration back to freshwaters to spawn. They are the anadromous fish species: the Atlantic salmon (*Salmo salar*) [Webb et al., 2007], the Sea trout (*Salmo trutta*) [Thorstad et al., 2016] or the Sea lamprey (*Petromyzon marinus*) [Silva et al., 2013] being among the best known species along the European West coasts. Other diadromous species have a reverse migration cycle, they are the catadromous fish: they spawn in marine environment and mature in freshwaters before performing their downstream migration back to the sea to spawn. One of the best known catadromous species is the European Eel (*Anguilla Anguilla*) [Tesch, 2007].

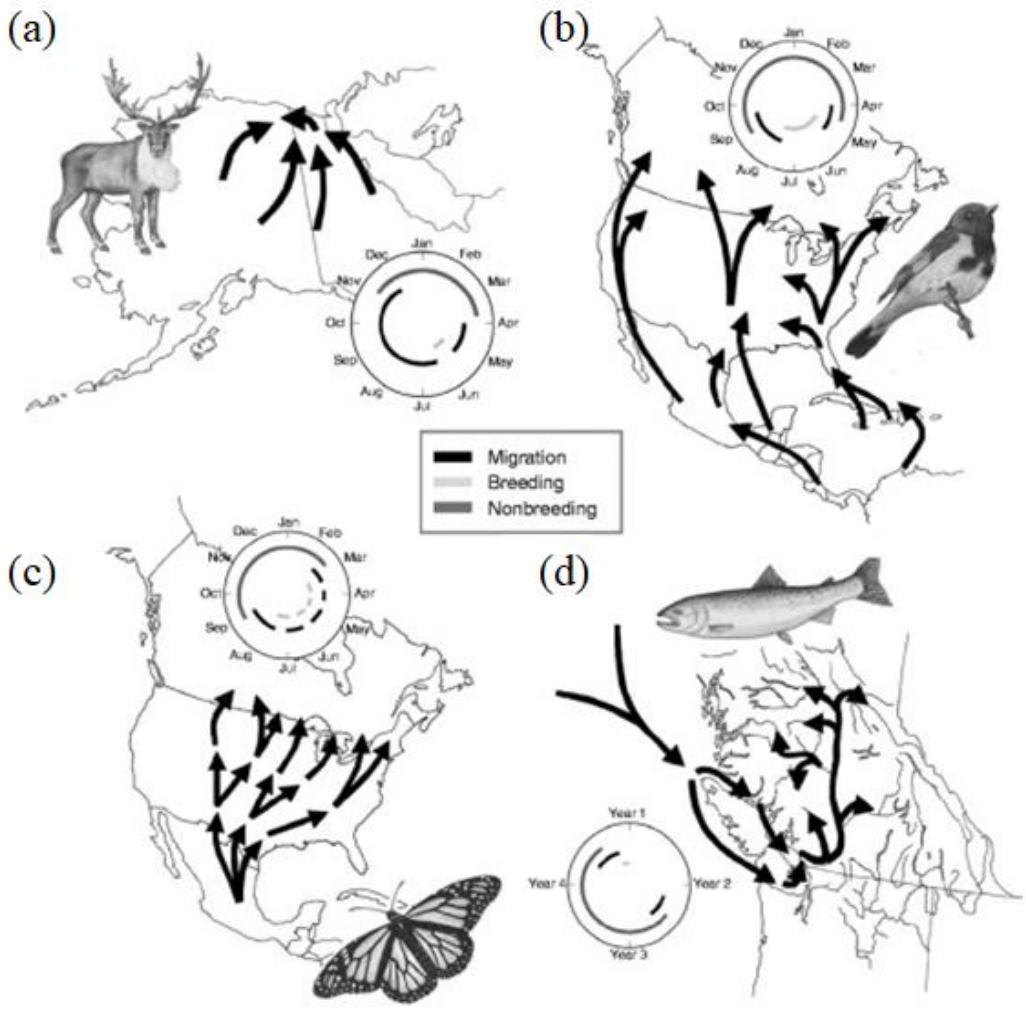


Figure 1 – Examples of migration patterns in different taxa: (a) the Porcupine caribou herd (*Rangifer tarandus*), (b) songbirds (American redstart *Setophaga ruticilla*), (c) insects (Monarch butterfly *Danaus plexippus*), and (d) fish (Pacific salmon *Oncorhynchus spp*). Migratory pathways depicted by arrows are only generalizations. Figure from [Hobson and Norris, 2008].

For all species, migratory movements are therefore vital as they ensure essential biologic functions, with the aim of completing their life cycle. However, migratory movements come with costs in time, energy, and risk that depend on the travelled distance [Bernatchez and Dodson, 1987].

1.2 Ecological continuum: rivers fragmentation and restoration policies

Fish ability to achieve their migration is therefore a key for the survival of the population. However, anthropic use of freshwater rivers, for example for energy purpose, tends to fragment

river courses and hamper fish migration pathways (Figure 2). The numerous facilities along the rivers are obstacles to the fish migration [Larinier, 2001] affecting the composition of diadromous fish communities, their abundance and their spatial distribution [Drouineau et al., 2018]. Facilities cause direct mortality, for example through turbines for a hydropower plant, or indirect mortality through migration delay, over-predation or overfishing, among others causes [Drouineau et al., 2018]. Direct and indirect consequences of habitat fragmentation contribute to the decline of the diadromous fish populations whose stock has dramatically decreased for several decades [Limburg and Waldman, 2009]. Both anadromous and catadromous species are concerned. Among them, Atlantic salmon and European eel have seen their population radically decline in French rivers and their distribution area reduce. The Atlantic salmon has disappeared from many French rivers for the two past centuries and maintains nowadays sustainable populations only in a few rivers of the North-West and South-West of France [Thibault, 1994, April et al., 2021] (Figure 3.a). The European eel is listed on the IUCN red list as a critically endangered species, its population stock has also drastically declined since the mid-1970s [Aalto et al., 2015, Dekker, 2016] (Figure 3.b). Looking beyond diadromous fishes, obstacles to river flow may impact the composition of the entire freshwater fish communities [Barbarossa et al., 2020].

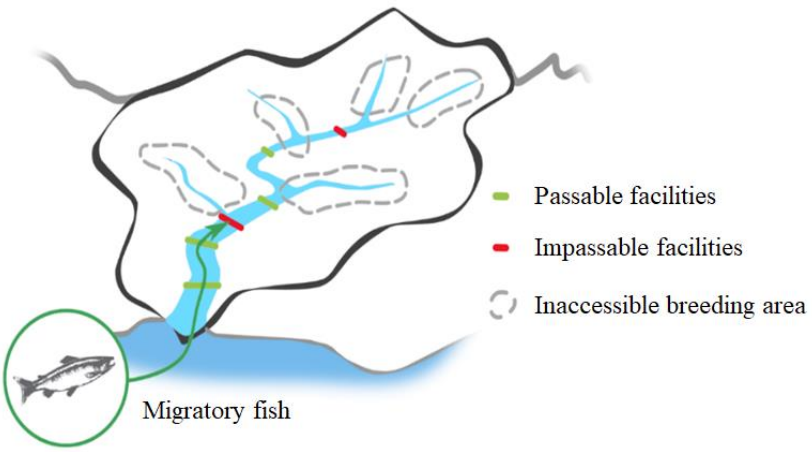


Figure 2 – Scheme of the fragmentation of a riverbed. The facilities passable by fish are displayed in green. The obstacles to upstream migration are displayed in red while the inaccessible breeding area induced are displayed in grey. Translated from French. © Office International de l'Eau.

To recover the ecological continuum, restoration policies and mitigation measures have been implemented (Water Framework Directive 2000/60/EC for water policy of the European Parliament and of the Council of the 23^{th} October 2000). Upstream and downstream migrations have to be ensured. Removing the sources of fragmentation is the most efficient solution to restore the river's continuity, however it is not always possible, nor desirable. Other restoration solutions, such as pool fish way [Larinier, 2002], natural bypass channel [Larinier, 2002], catch

and transport, fish lift [Travade and Larinier, 2002] ensure the crossing of facilities by fish during their upstream migration. Regarding downstream migration, different solutions have been designed such as bypass associated with guiding devices (such as rack) or fish-friendly intake. But regarding the large hydropower plant, the only solutions available today are operational actions planned to open passage for fish. This is the case of turbine management for hydropower plants [Teichert et al., 2020, Song et al., 2019]. Fish-friendly turbines can also be installed [Hogan et al., 2014]. As the first producer of hydroelectricity in France, Électricité de France (EDF) is an important actor of the river continuum in France. EDF must implement those mitigation measures to meet the policies' requirements and preserve the aquatic biodiversity while optimising production costs.

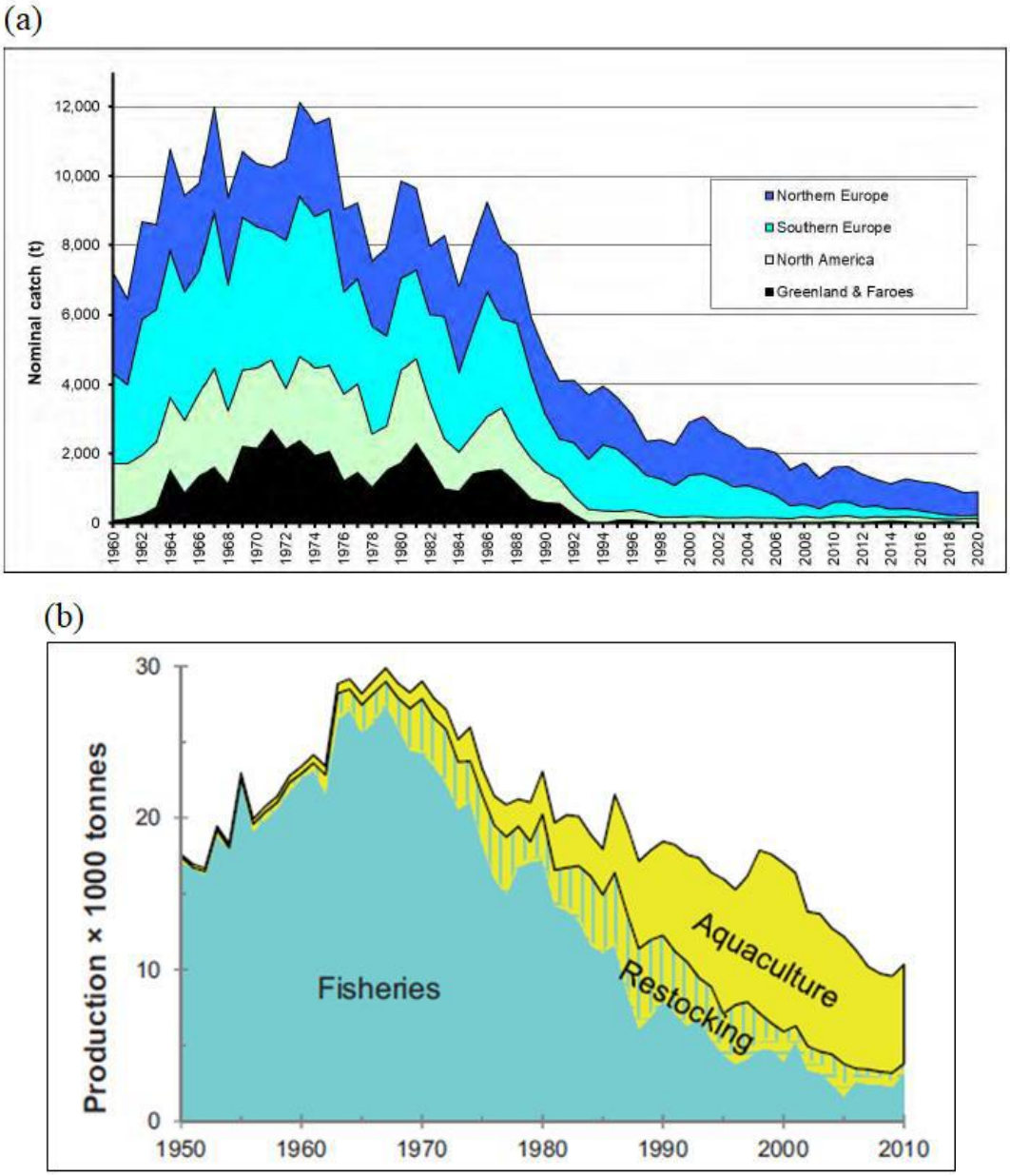


Figure 3 – (a) Total reported of nominal catches of Atlantic salmon (tonnes round fresh weight) in four North Atlantic regions between 1960 and 2020, from [April et al., 2021]. (b) Time trend in the European eel production, combining fishing yield from the wild stock with aquaculture (using wild glass eel), from [Dekker, 2016].

2 Acoustic imaging: a non-intrusive tool to monitor fish?

2.1 Fish monitoring: an overview of the existing methods

To assess the status of migratory fish populations, ensure their proper management or evaluate the effectiveness of measures to restore ecological continuity, the monitoring of fish migration is essential. The quantitative and qualitative analysis of the movements of different species according to environmental conditions provide information for the design of fish passes and for

the adaptation of operational management of facilities. Multiple techniques and tools are used to retrieve this valuable information. Some of them require to catch fish individuals, others record their movements without any handling (Figure 4).

The most common capture-dependent methods are recreational and commercial fisheries and scientific surveys [Clay et al., 1998, Bohlin et al., 1989]. Catching fish enables to get punctual information about their position, to measure their morphological characteristics [Durif and Elie, 2008] and to collect samples providing other life history traits information through genetic or sclerochronology methods. After their capture, fish migration behaviour may be monitored using telemetry methods [Thorstad et al., 2013, Brownscombe et al., 2019]. Tags are inserted in or fixed on the fish body. Fish are then released and tracked in the environment as long as receivers are present nearby their successive locations. Telemetry is a widely used method for measuring efficiency and/ or optimising fish passage solutions [Szabo-Meszaros et al., 2019, Jebria et al., 2021, Larinier et al., 2005].

However, despite the valuable information that capture-dependent methods provide, they present some limitations in their use. The number of fish that can be monitored is limited. Fisheries and scientific surveys are dependent on the environmental conditions and only provide punctual monitoring over time. Telemetry is limited to the number of tags available as well as the time and constraints inherent to the required fish manipulation. Nevertheless, the main disadvantage of capture-dependent methods is their intrusiveness. The injection of the tag may have an impact on the fish behaviour as they can increase their level of stress [Ramstad and Woody, 2003]. Overall, capture-dependent methods are expensive to implement.

The capture-independent methods, non-intrusive, are also widely used to monitor fish migration. Visual observation through direct observation or through optical cameras enable to collect information on fish passages [Cappo et al., 2006, Wilson et al., 2014]. Optical cameras capture on screen fish passages without any manipulation. These cameras must be installed at a narrow passage location where a light illuminates the section of water. Hence, optical cameras can record fish passages at a very limited range and visibility largely decreases in turbid waters and with limited light conditions. They are thus usually limited to fish ways installation [Travade and Larinier, 2002]. Fish counter can also be used to count fish passages, such as resistivity fish counter [Lethlean, 1954, Dunkley and Shearer, 2006]. Other techniques are under development such as quantitative environmental DNA (eDNA) [Thomsen and Willerslev, 2014]. eDNA is the genetic material extracted from the studied environment corresponding to the river's water in case of fish presence assessment. eDNA analysis makes it possible to attest the presence of a distinct aquatic species but is not able yet to quantify its population abundance [Rourke et al., 2021, Vautier et al., 2023].

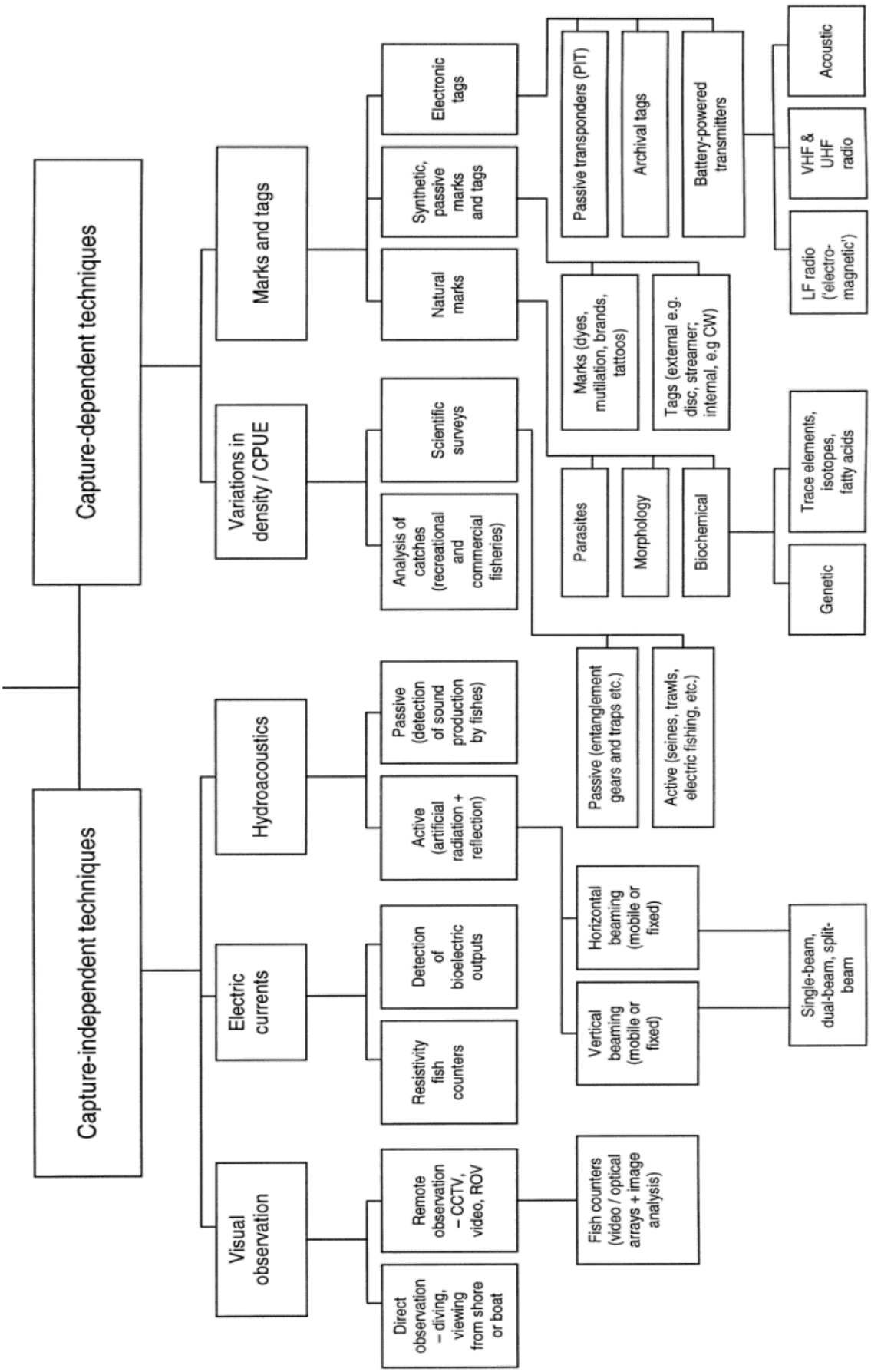


Figure 4 – Classification of fish monitoring methods with distinction between capture-independent techniques and capture-dependent techniques, from [Lucas and Baras, 2000].

Active acoustic devices, or sonars, are another well-known non-intrusive method. At the opposite of passive acoustic devices that monitor wildlife and environments using sound recorders, sonars emit acoustic pings while receiving back the echoes reflected by all the objects present in their beams. They have been widely used for several decades in freshwater and marine studies [Simmonds and Maclennan, 2005, Rudstam et al., 2012].

2.2 Acoustic imaging

Since the development of the first types of sonar, many technological evolutions have been made (Figure 5), improving the capacity of the sonars to collect information [Martignac et al., 2015]. Among the last technological evolutions, high-frequency multi-beam sonars, called Acoustic Cameras (AC), register video-like data. Each acoustic image, or frame, of an acoustic video (Figure 6) is a two-dimension representation of all the echoes reflected by the objects that the acoustic signals of each beam, meet and receive. Consequently, the frame can be seen as a top-view image of the volume on which the AC emits the acoustic signal and records the consecutive echoes. The AC range depends on the acquisition settings and on the AC types; particularities of the AC will be detailed in Chapter 1. Thanks to its very high frequency, the AC acoustic signal is reflected by every part of the fish body, and each detected echo is displayed on the acoustic frame by a level of brightness that is proportional to its intensity. Thus, the acoustic videos allow a direct observation of the fish morphology, from its head to the end of its tail, as well as its movement and behaviour. Consequently, acoustic imaging provides a new way to discriminate fish from other objects compared to former sonars echograms [Pratt et al., 2021] making them promising tools for fish monitoring [Martignac et al., 2015] even in turbid waters [Staines et al., 2022]. However, fish species identification is applicable with confidence only on a small number of species with large sizes and distinctive morphological features [Jones et al., 2021, Wei et al., 2022].

AC have first been designed for military applications such as intruder detection, before being used for monitoring submerged structures. It was only later that their use was diverted to be applied on fish monitoring. AC are today used for many different fish monitoring applications such as: abundance estimation [Hayes et al., 2015, Mora et al., 2018, Artero et al., 2021, Sibley et al., 2023], migration behaviour [Kirk et al., 2015, Capoccioni et al., 2019, Lenihan et al., 2019, Lenihan et al., 2020, Keeken et al., 2020], movement within habitats [Lagarde et al., 2021, Bennett et al., 2020, Staines et al., 2022] and predation and antipredator behaviours [Becker et al., 2013, Price et al., 2013].

Thanks to their characteristics, AC are more and more used to count fish [Holmes et al., 2006, Lagarde et al., 2020] and to measure and describe their morphology. Hence, despite potential

errors and operator bias [Burwen et al., 2010, Daroux et al., 2019, Helminen et al., 2020], length measurement on the acoustic frames is accurate and is an interesting input of AC [Grote et al., 2014, Gurney et al., 2014, Zhang et al., 2014, Lin et al., 2016] providing relevant information to identify species, evaluate age composition and estimate fish biomass [Wei et al., 2022].

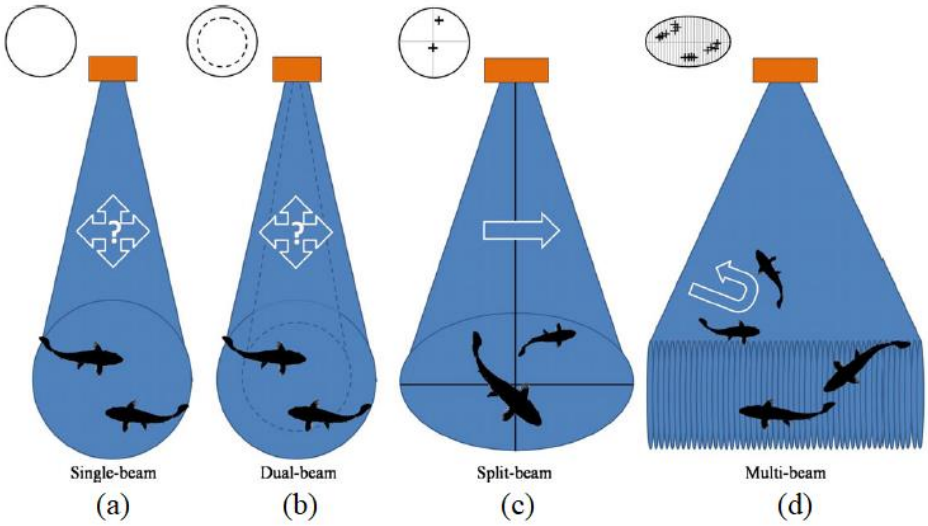


Figure 5 – Scheme of the different technological evolutions of the sonars, from the former sonars (single-beam (a), dual-beam (b), split-beam (c)) to the multi-beam sonars (d), from [Martignac et al., 2015].

Even if AC present a lot of advantages for fish migration monitoring, they still have limits to overcome to be fully exploited. In terms of operational requirement, AC must be connected to a power supply and to a computer. Besides, AC location must be carefully chosen because of its limited covered volume in order to maximise fish recording [Martignac et al., 2015]. Data storage and data processing are also two of the main constraints of AC: in the aim of long-term monitoring, the analysis of the high amount of data recorded is very time-consuming for an operator. Besides, operator effect on the analysis is an important element to consider. Many studies discuss the potential bias an operator can bring during its acoustic videos visualisation, regarding the counting [Keefer et al., 2017, Lagarde et al., 2020, Braga et al., 2022] or the fish body length measurement [Hightower et al., 2013, Grote et al., 2014, Daroux et al., 2019].

The development of an unbiased automatic method to analyse acoustic imaging through species counting, identification and behaviour analysis is thus a key step to reach in order to extract all the information available on AC data.

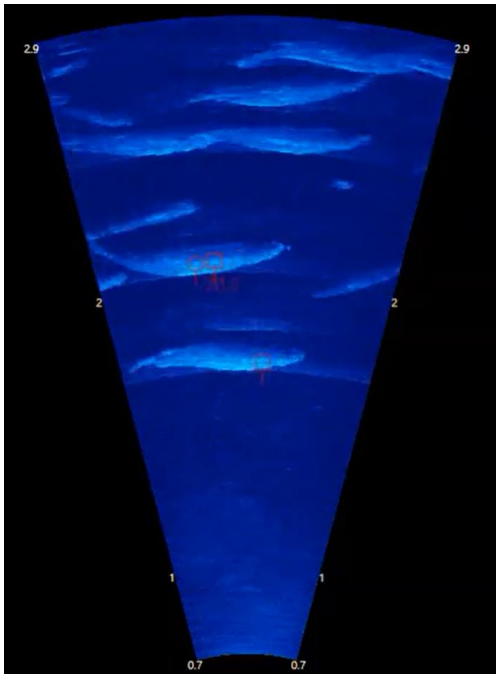


Figure 6 – Acoustic camera frame capturing salmon passages by RSC/Reeds Sonar Consultancy LTD. The video has been recorded using an ARIS Explorer 3000 from 0.7 to 2.9 meters.

2.3 Acoustic imaging analysis: an overview

The analysis of acoustic videos (Figure 7) is a challenge that is currently addressed in the literature. As AC record successive frames building true videos, [Mueller et al., 2008] and [Boswell et al., 2008] have been the firsts to offer a computer-vision based pipeline to analyse them as optical videos. The semi-automated pipeline of [Mueller et al., 2008] has been designed for American eel's (*Anguilla rostrata*) identification. It relies on the extraction of shape features and pixel features at each eel detection followed by a classification for which three distinct classifiers (Discriminant function analysis, K-nearest neighbours, Neural Network) have been tested [Duda et al., 2001]. In their study, [Boswell et al., 2008] introduced a pipeline to detect and track fish passages from acoustic videos in order to provide an abundance estimation. Both developed pipelines have the same scheme of analysis based on image-processing tools: a first part focuses on enhancing the fish images, a second part aims to detect and track moving objects from frame to frame. Their characteristics of interest, such as fish morphology or motion characteristics, are then calculated depending on the method's objectives.

Following these pioneer approaches, other pipelines have been developed. From fish counting and sizing, to fish identification up to species classification, each of these pipelines addressed the analysis required for operational needs, such as fish population monitoring, with their own approaches and case studies. As a first step to analysis, different semi-automatic or fully automatic fish counting methods, and for

some of them sizing methods as well, have been designed using the software EchoView [Kang, 2011, Helminen and Linnansaari, 2021]; or using implemented computer-vision methods [Han et al., 2009, Bothmann et al., 2016, Shen et al., 2023]. To go further in the acoustic videos analysis, [Kandimalla et al., 2022] and [Fernandez Garcia et al., 2023] developed a method of fish identification based on deep learning algorithms: the aim was to distinguish fish from debris or other non-fish objects. It can be considered as a first step to focus on the passage of interest prior to species classification. [Connolly et al., 2022] achieved similar objectives by applying deep learning classification on both fish images and fish shadow images.

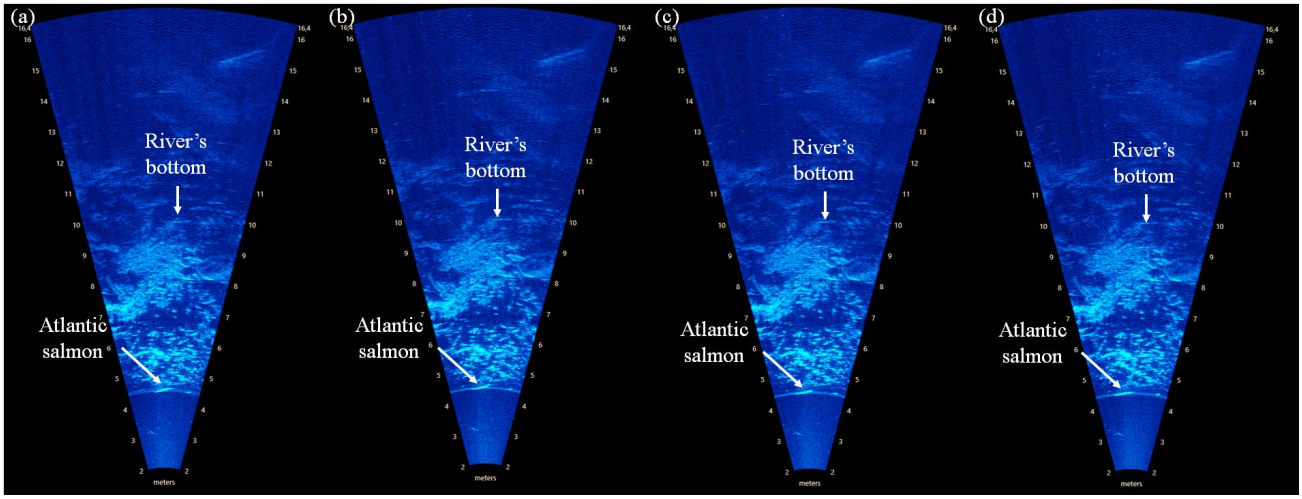


Figure 7 – Examples of four successive frames of the passage of an Atlantic Salmon captured on an ARIS Explorer 3000 acoustic camera at low-frequency resolution (1.8MHz).

Hence, automatic fish species identification is another key step to fully exploit the acoustic imaging capacities. According to their specific morphology, most studies focus on the discrimination of eels from debris or other non-anguilliform fish, through morphological analysis [Mueller et al., 2008, Bothmann et al., 2016], or deep learning classification [Yin et al., 2020, Zang et al., 2021]. Other studies proposed species classification through templates matching, using acoustic shadows [Langkau et al., 2012] or through high-resolution images [Yu et al., 2016]. Both methods require specific site configurations. On the one hand, recording acoustic shadows requires that the beam of the AC hits a defined surface such as a projection plate or the bottom of the river. On the other hand, retrieving high quality images with detailed fish aspects requires to limit the AC range to two meters [Yu et al., 2016]. Although allowing to get fish images with fin and aspect body details, the recording constraints are important and do not seem to be applicable on fish migration monitoring for which an operator wants to cover the widest stretch of the river.

Species classification from AC data is therefore challenging due to the nature of the frames. On optical images, a certain degree of detail is available for species classification: colour, number

and position of the tails, body aspect (Figure 8) and swimming locomotion, i.e., how the fish deforms its body to move. On AC data, many of these details are not available. Colour cannot be retrieved as the frames are reconstructed from echoes intensity. Only the swimming locomotion and the global shape of the fish can be visualised, with no details on the body aspect or on the pectoral and dorsal fins because of their low thickness and the poor resolution of acoustic frames.

Figure 8 – Image of an Atlantic salmon captured by an optical camera at a fish way location. © Fédération de pêche de la Somme.

2.4 Fish locomotion and morphology: useful tools for species classification on acoustic imaging?

Fish locomotion corresponds to the body deformation of the fish due to the physical forces that apply on him in a water environment [Lindsey, 1978]. It is used to classify species depending on their kinematic characteristics. [Lindsey, 1978], [Sfakiotakis et al., 1999] and [Webb, 1984b] review the different types of swimming mode: on one side, fish with Body and/or Caudal Fin (BCF) propulsion, with lateral body undulations running from head to tail; on the other side fish with Median and/or Paired Fin (MPF) propulsion [Lindsey, 1978] (Figure 9). Fish may switch from one type of propulsion to the other depending on their behaviour. Foraging usually corresponds to MPF modes as it allows fish a greater stability and manoeuvrability [Sfakiotakis et al., 1999]. Conversely, for high speed or acceleration, fish usually exhibit BCF modes, more suitable for this kind of behaviour [Lindsey, 1978].

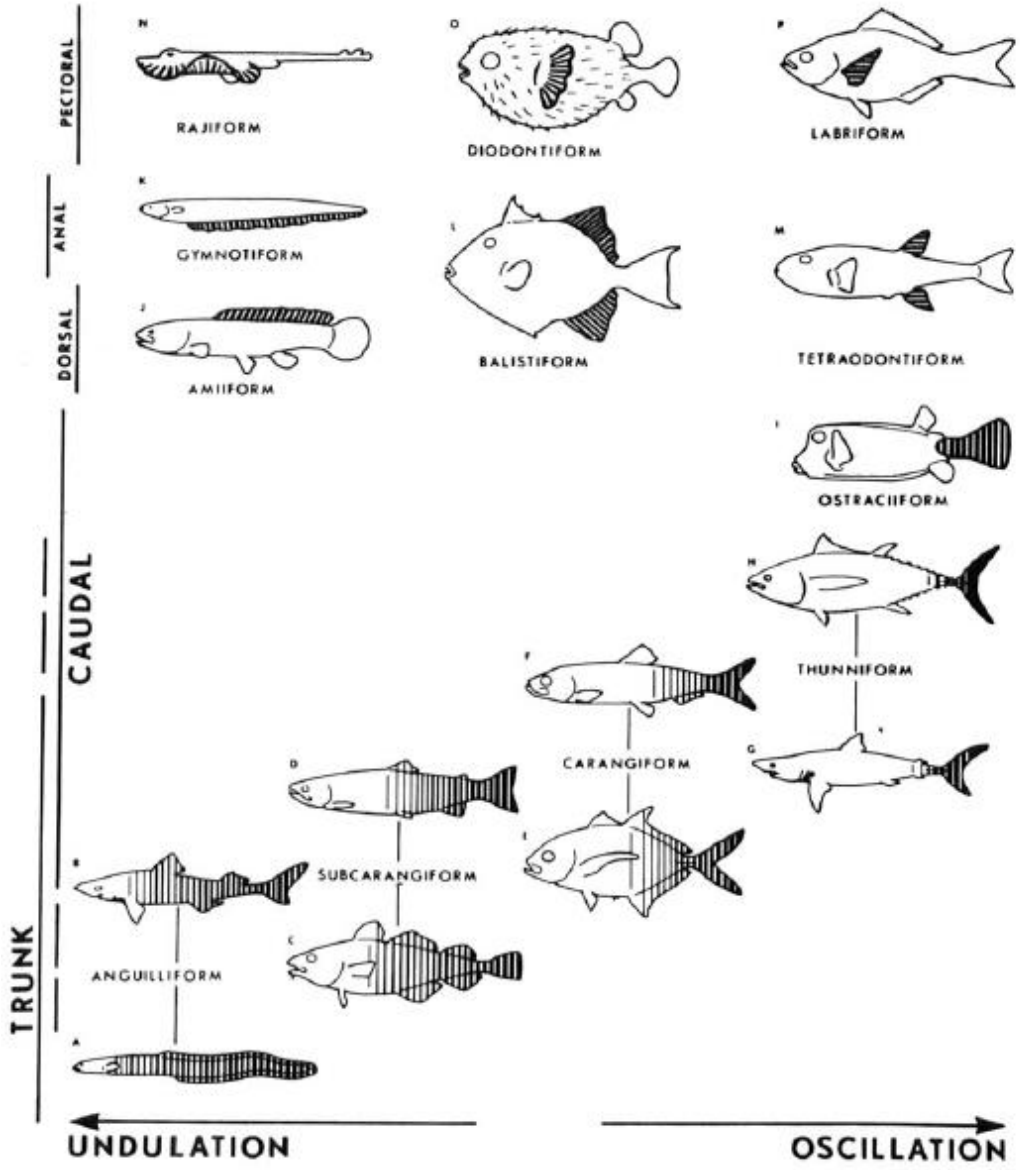


Figure 9 – Scheme of fish swimming modes. The vertical axis shows the propulsive contributions of body and fins while the horizontal axis displays swimming mode from serpentine undulation to oscillation, from [Lindsey, 1978].

Depending on the fish orientation to the AC, MPF might be harder to identify on acoustic imaging than BCP propulsion and thus will not be detailed here. Four main swimming BCF modes have been classified by [Breder and Society, 1926] and [Lindsey, 1978]: the anguilliform, the subcarangiform, the carangiform and the thunniform modes (Figure 10). The wavelength and the amplitude of the propulsive wave are the two main characteristics varying between each of those four modes [Sfakiotakis et al., 1999]. This classification does not consider any taxonomic affinities, it only relies on the swimming locomotion of fish and on some morphological considerations. Hence, fish from different taxonomic groups can therefore correspond to the same swimming mode [Lindsey, 1978], such as the European eel from the Anguillidae family and the European catfish from the Siluridae family, which both have an anguilliform swimming

mode. The swimming mode may also evolve during the fish life cycle, especially between its young and adult states. Many fish exhibit an anguilliform mode during their young state before evolving [Lindsey, 1978]. If swimming locomotion modes have been classified in the four classes described above, [Di Santo et al., 2021] highlight potential overlapping between them with a lot of individual and intra-specific variability in parameters such as wavelength, and an overlapping in oscillation amplitude between swimming modes. Besides, as highlighted in the review of [Liao, 2007], the fish locomotion knowledge mainly comes from studies conducted in still water or steady flows. Their behaviour in running water should be investigated as well as its impacts on the swimming motions of fish.

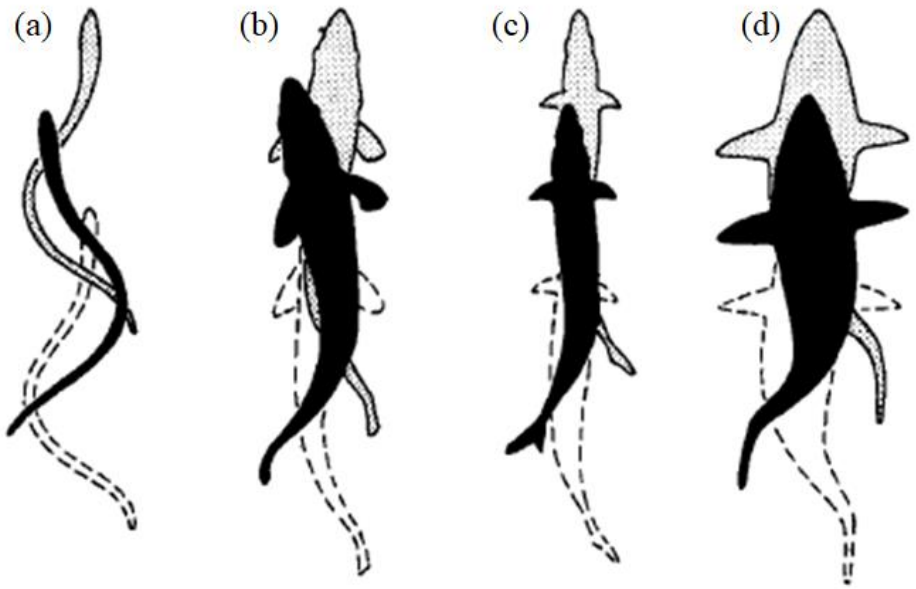


Figure 10 – Scheme of the four swimming BCF modes classified by [Breder and Society, 1926] and [Lindsey, 1978]: the anguilliform swimming mode (a), the subcarangiform swimming mode (b), the carangiform swimming mode (c) and the thunniform swimming mode (d), from [Lindsey, 1978].

In addition to a panel of swimming modes, freshwater fish species also display a large diversity of morphologies, their global shape and length vary depending on the species. Fish global shape relationship with fish locomotion has already been discussed in the literature [Webb, 1984a, Webb, 1984b, Webb, 1988]. [Tytell et al., 2010] highlighted that they are both highly correlated. Significant differences between some fish having BCF propulsion and others having MPF propulsion have been found by [Friedman et al., 2021] for marine species, despite existing overlap. The shape of fish with BCF propulsion has been observed to be more streamlined and have a narrower caudal peduncle than MPF ones [Friedman et al., 2021]. Among the BCF propulsion fish, elongate fish species usually swim in an anguilliform mode, others tend to swim in subcarangiform or carangiform modes [Tytell et al., 2010]. In addition to body shape, other morphological characteristics are used for species identification, the main ones being

area, length and width [Strachan et al., 1990, Pornpanomchai et al., 2013, Li and Hong, 2014]. Among them, fish length can be considered as one of the main criteria for fish species identification in AC data [Martignac et al., 2015, Daroux et al., 2019].

Fish locomotion and fish morphology may therefore provide useful information for species identification. Automatically extracting that information from acoustic videos requires to apply computer-vision methods initially developed for optical video analysis and to adapt them to the AC data properties and resolution.

3 PhD's objective

To the best of our knowledge, no studies have yet investigated the potential of a single analysis pipeline for different types of AC as well as for different monitoring sites with their own recording constraints and proper fish population morphological characteristics and diversity. This PhD objective is to develop a single operational real-time and cross-camera solution (Figure 11) to answer our two main questions in order to overcome some of the main technological locks of the acoustic imaging:

- How to extract data of interest, through fish detection and tracking, among the high amount of acoustic images?
- How to classify fish species from acoustic imaging?

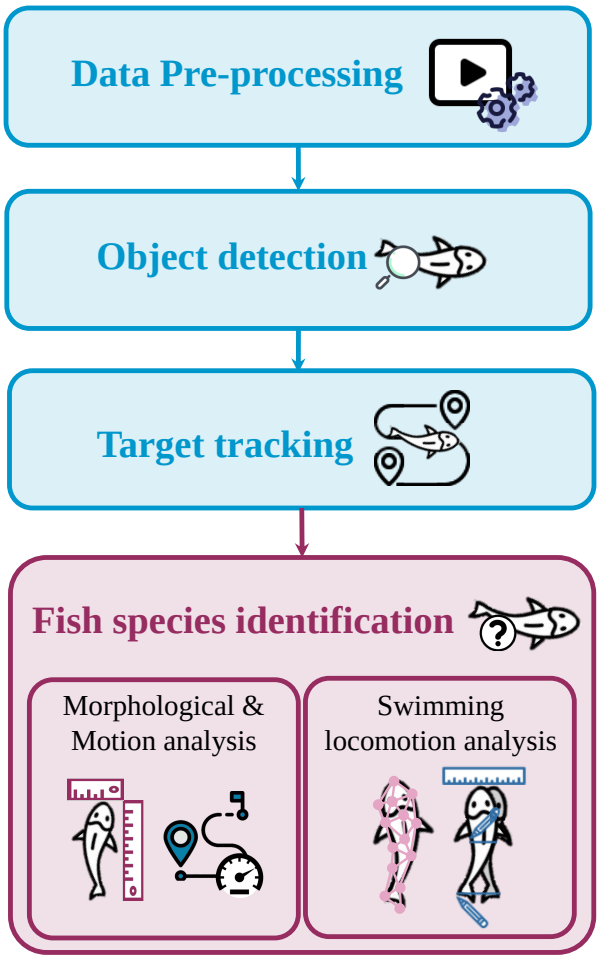


Figure 11 – The global pipeline is composed of a first part of data pre-processing followed by fish detection and tracking. The last step is the fish species identification for which morphological and motion analysis is carried out on one side and swimming locomotion analysis is carried out on the other side.

3.1 Detect and track: how to sort data of interest among the high amount of recorded AC data?

The first part of the pipeline aims to successfully filter the high amount of AC data by extracting only the one of interest (Figure 11, in blue). Based on the concept of the video-echogram already explored by [Mueller et al., 2008], the method proposes a first automatic and operational analysis of the videos to extract the targets passages putting aside the passages with no data of interest. Extracting moving objects and tracking them are two essential steps that have therefore been widely used and evaluated in pipelines for acoustic videos analyses [Mueller et al., 2008, Boswell et al., 2008, Han et al., 2009, Bothmann et al., 2016]. However, in the proposed pipeline, the image processing is pushed a little further to integrate the high level of noise of the AC images [Kandimalla et al., 2022] as well as to address the issue of fragmentations of the fish body echoes [Mueller et al., 2008, Han et al., 2009]. We thus propose a novel approach of image restoration to balance the impact of noisy images on the next analysis

steps. These image processing steps will extract the data of interest but also will pre-process them for species identification.

3.2 Identification: how to carry out species classification from the AC data?

As mentioned previously, the information that can be extracted from acoustic videos to allow species identification is quite limited. This PhD work focuses on two main categories of characteristics that can support the fish species identification on acoustic videos:

- their morphological and motion, such as velocity and trajectory, characteristics; and,
- their swimming locomotion characteristics, i.e, how the fish deforms its body to swim.

Size and shape are the main characteristics for species discrimination [Able et al., 2014] but are often not clearly recorded in the acoustic videos [Parsons et al., 2017]. Morphological features are therefore usually combined with motion information [Wei et al., 2022] as in [Mueller et al., 2008] and [Bothmann et al., 2016]. Both studies have proposed semi-automatic or fully automatic acoustic imaging analysis based on computer vision methods to identify European eel (*Anguilla Anguilla*) among non-anguilliform species. Despite reaching high performance, [Mueller et al., 2008] do not propose a fully automatic analysis, its method requiring successive manipulations with different software. [Bothmann et al., 2016] provides a real-time and automatic pipeline. However, their data have been recording until 6 meters, providing optimum quality images barely obtained during *in situ* monitoring studies. Thus, they have not been confronted with issues of image fragmentation, encountered with low quality images [Mueller et al., 2008, Han et al., 2009].

- **Task 1: Morphological and motion analysis.**
  Based on the works of [Mueller et al., 2008] and [Bothmann et al., 2016], I first focus on morphological and motion characteristics to propose an operational and generalist method by considering data recorded at higher range and therefore with lower resolution (Figure 11, on the left, in pink). The method is tested on data recorded by two types of AC, with different characteristics providing different covered volume, and on two different monitoring sites, with their own populations of European eels which morphologically differ. Besides, the method aims to fully exploit the video dynamic in the morphological analysis by studying the evolution of the morphological characteristics along the target movement. Hence, it will allow us to consider how the position and angle of the target relative to the AC transducer influence the target's aspect [Parsons et al., 2017].
- **Task 2: Swimming locomotion analysis.**
  Morphological and motion analysis are not the only characteristics that enable an oper-

ator to identify fish species when visualising the acoustic videos. The deformation of the fish body along its movement, i.e., its swimming locomotion, is also a key characteristic for identification. Hence, when looking at a fish image recorded on AC video (Figure 7.a), it is difficult, or even impossible for an operator to identify species. The successive images of the same fish (Figure 7.b-d) allows the operator to give an identification result. Characterising this swimming locomotion requires to track and quantify the body deformation at each detection of the target, using deformable models (Figure 11, on the right in pink). To our knowledge, such deformation analysis has not been yet explored on acoustic imaging. The discrimination ability of this kind of information will be investigated as a complementary characteristic to enable multi-species classification from AC data.

3.3 Organisation of the manuscript

The PhD manuscript is organised in four chapters aiming to reflect the thinking diagram that drives my work (Figure 11). After the presentation of our datasets in Chapter 1, the Chapters 2, 3 and 4 describe the methods we developed to reach the two objectives of the PhD.

- The Chapter 1 introduces the acoustic imaging and presents the different types of AC, the monitoring sites and how the data used in this PhD have been recorded. The different datasets are then described in detail.
- The Chapter 2 focuses on the handling of the high amount of acoustic videos by detecting and tracking targets of interest.
- The Chapter 3 introduces the analysis of the fish morphological and motion characteristics as one of the ways to carry out species identification through the pipeline published in [Le Quinio et al., 2023]. An application to an operational study case is then given.
- The Chapter 4 focuses on fish locomotion as a second source of information available to identify species from AC.
- The PhD manuscript will end with a general discussion underlying the main conclusion of my works, the scopes of improvement and the perspectives that can be pursued following this PhD.

Chapter 1. Material and methods

Contents

La vision est l'art de voir les choses invisibles.

Jonathan Swift

1 Acoustic imaging

1.1 General principles

Acoustic cameras (AC) are high frequency multibeam sonars designed to create high resolution images [Martignac et al., 2015]. The multibeam sonars correspond to a succession of n single beams that covers a 3D water section (Figure 12.a) defined as their Field Of View (FOV). When the acoustic pulse encounters a target, an echo is sent back to the AC. The distance between the camera and the target is calculated from the delay between the acoustic pulse emission and the reception of the target's echo. Each single beam is subdivided in multiple segments of equal distance called "samples" along the range axis. The target position is therefore recorded on this axis, but not along the depth and width axis. The entire target morphology as well as its position in the FOV are therefore apparent on AC frames thanks to all of the echoes recorded in the contiguous single beams.

Hence, echoes from the n adjacent beams will together map the reverberation of the AC FOV creating the 2D acoustic images (Figure 12.b) [Belcher et al., 2001] as a top view of the covered water volume. The total FOV is consequently divided in multiple cells, whose horizontal dimension is defined by the width of a single beam and vertical dimension is defined by the height of a sample. The vertical dimension of cells is defined as the down-range resolution $r_{samples}$, expressed in millimetres, and corresponds to the ratio of the window length by the number of samples (eq. 1) [Burwen et al., 2010]. The window length corresponds to the distance, in meters, covered by the AC, i.e., the distance between the maximum and the minimum range of the AC FOV. The horizontal dimension of the cell is defined as the cross range resolution r_{beams}, that relies on the FOV opening and on the number of single beams. r_{beams} varies with the distance to the AC: it decreases when this distance increases. r_{beams} can be calculated from geometric formula using (eq. 2).

$$r_{samples} = \frac{w_L \cdot 1000}{N_{samples}} \qquad \text{(eq. 1)}$$

$$r_{beams} = \frac{\frac{\theta \pi R}{180} \cdot 1000}{N_{beams}} \qquad \text{(eq. 2)}$$

With $r_{samples}$, the down-range resolution, in millimetres; r_{beams}, the cross range resolution, in millimetres; w_L, the window length of the AC, in meters; $N_{samples}$, the number of samples of the AC; θ, the opening of the FOV in degrees; R, the range of the sample, in meters; and, N_{beams}, the AC number of beams.

The information displayed on each cell of the acoustic images corresponds to a proxy of the energy reflected by the targets and recorded by the AC. Hence, AC are not designed to measure the acoustic properties of the fish echoes as recorded by former sonars

[Simmonds and Maclennan, 2005] because the high frequencies used reflect on the whole body of the fish. The higher the reflecting energy is, the lighter the cell will be displayed on acoustic images.

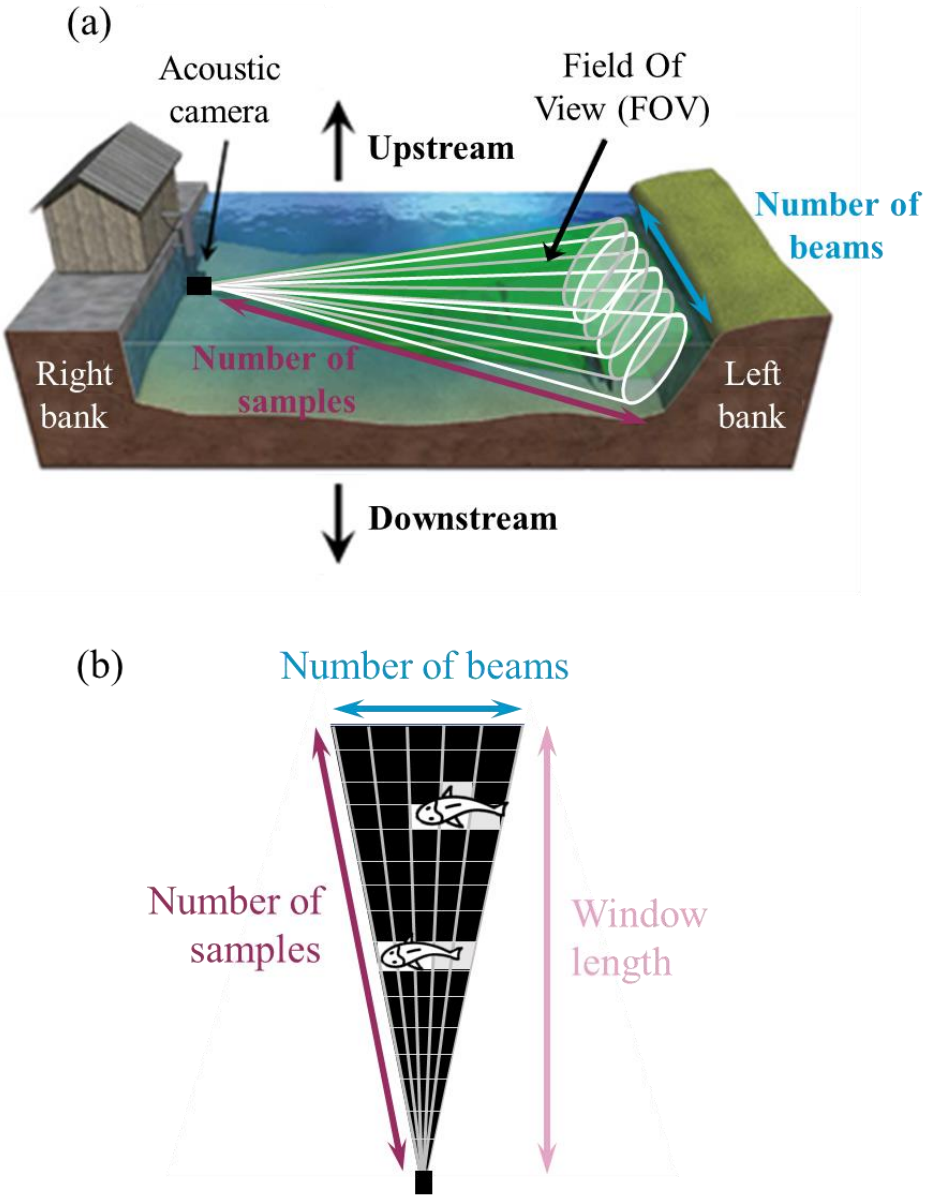


Figure 12 – (a) Scheme of the installation of an acoustic camera on the right bank of a river. The covered volume of the camera is represented in green. The multiple beams of the AC are drawn in white. Adapted from [Guillard and Lebourges-Dhaussy, 2014]. (b) Scheme of an acoustic image. The samples are represented as horizontal grey lines while the beams correspond to the vertical ones. The intersection between a sample and a beam forms a cell. The height of the image relies on the window length.

Acoustic videos are therefore built from the successive acoustic images recorded by the AC. The number of acoustic images per second defines the frame rate of the acoustic videos.

1.2 AC images and videos

Acoustic images are defined as a grid of cells that depends on a number of beams and on a number of samples. However, a classic image is expressed as an array of pixels. To each pixel is assigned a value called pixel intensity. In this PhD manuscript, in contrast to the acoustic images defined by cells, the AC images are defined by pixels. AC images are arrays of pixels (Figure 13.a) that correspond to the translation of the acoustic images cells into classic images (Figure 13.b).

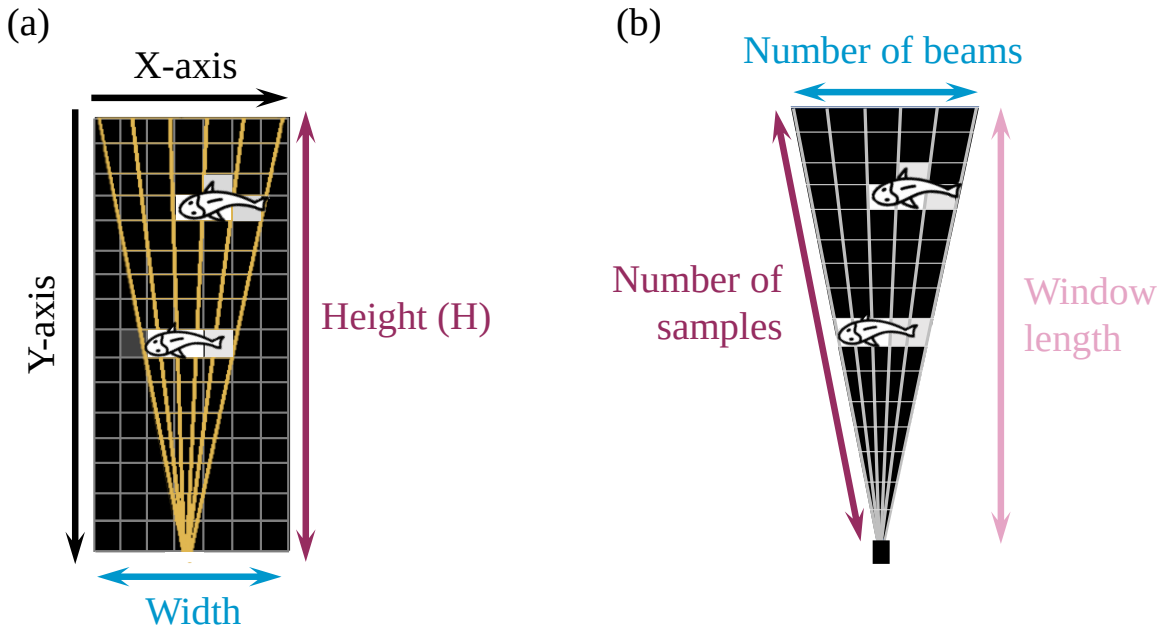


Figure 13 – (a) Similar scheme as in Figure 12.a shows an acoustic image with the samples and beams displayed in grey. (b) Scheme of an AC image. It corresponds to the translation of the acoustic image (a) defined by samples and beams (in yellow) into a classic image corresponding to a grid of pixels (in grey). The X-axis and Y-axis are defined as the image's width and height respectively.

The energy reflected on the AC transducer and displayed on the acoustic images is converted in grayscale pixel intensity on the AC images. It means that the pixels values range from 0, the darkest, to 255, the lightest pixel intensity. AC images are noisy images, i.e., they display random variation of the pixel intensities. The noise present on the AC images is called salt-and-pepper noise. It is similar to sprinkling white and black dots on the image [Boncelet, 2009]. AC images are defined by their height and width that are proportional to the number of single beams, the angular opening of those beams, the number of samples and the window length.

As opposed to the acoustic image resolution that corresponds to the resolution of one beam or one sample, the pixel resolution corresponds to the resolution r of one pixel of the AC images. The pixels of AC images being squares, the resolution is the same along the X-axis and the

Y-axis. Pixel resolution is proportional to the acoustic image resolution and can be calculated through (eq. 3).

$$r = \frac{r_{samples} \cdot N_{samples}}{H} \quad \text{(eq. 3)}$$

With r, the pixel resolution, in millimetres; $r_{samples}$, the down-range resolution, in millimetres; $N_{samples}$, the number of samples of the AC; and, H, the height of the AC image, in pixels (Figure 13.b).

Pixel resolution therefore does not rely on the pixel range and has the same value no matter its position in the AC images. However, pixel resolution depends on the size of the AC image that is set up by the AC itself.

1.3 The acoustic camera devices

Several AC devices from different manufacturers are available. The first device to be commercially available was the Dual-frequency Identification Sonar (DIDSON) designed by Sound Metrics Corp. (Bellevue, WA, USA). DIDSON is operating up to a 1.8 MHz frequency with an opening of around 30° along the X-axis and of around 15° along the Z-axis with a maximum of 96 beams. The maximal window length of the DIDSON is 10 meters with a 1.8 MHz frequency. An improved device was later released, the Adaptive Resolution Imaging Sonar (ARIS, Sound Metrics Corp). Depending on devices, ARIS provides similar or better resolution with a number of beams reaching up to 128, and higher operating frequencies, up to 3 MHz for the ARIS Explorer 3000. Furthermore, ARIS window length is higher than DIDSON's at the same frequency (up to 15 meters with a 1.8 MHz). Both DIDSON and ARIS devices have a similar FOV opening around 30°. However, new AC with wider opening FOV have been recently designed. The BlueView cameras produced by Teledyne Technologies Inc. (Thousand Oaks, CA, USA) offer operating frequencies up to 2.25 MHz. Its FOV is composed of a maximum of 768 beams and has a 130° opening along the X-axis and up to 20° opening along the Z-axis. The last AC designed is the Oculus camera produced by Blueprint Subsea (Ulverston, Low Wood, UK). Some Oculus devices offer the same FOV opening than BlueView but lower operating frequencies of 1.2 MHz maximum with this FOV opening. In addition to provide wider FOV, the BlueView and Oculus cameras do not require maintenance. Hence, ARIS and DIDSON cameras are composed of a lens system where water pass by [Martignac et al., 2015]. The lens box is progressively lined with silt and therefore need to be regularly cleaned by an operator.

In our study, we use data recorded by three of these AC: ARIS Explorer 1800 (ARIS 1800), ARIS Explorer 3000 (ARIS 3000) and BlueView M900-2250-130 2D (BV) whose characteristics are summarised in Table 1. An example of the ARIS 3000 recorded image is displayed in Figure 14.

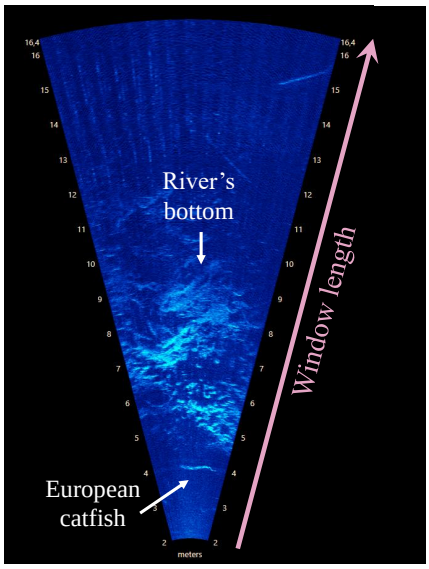


Figure 14 – Screenshot of an ARIS 3000 frame recording at the Ducey monitoring site. The window length starts at 2 meters up to 16.4 meters. A European catfish (around 1.1 meters long) can be observed at about 4 meters from the camera. The echoes observed from 5 meters to 12 meters correspond to the bottom of the river.

All of the three AC used provide Low-Frequency (LF) and High-Frequency (HF) modes for which specific frequency is available. On the one hand, the low-frequency mode enables to cover a larger range (Table 1) but with much lower resolution. On the other hand, the high frequency mode provides higher resolution but covers a more limited water volume, with a maximum range of 5 meters for the ARIS 3000, 15 meters for the ARIS 1800 and 10 meters for the BV. In my study, because the frame resolution was prioritised, the use of the high frequency modes for ARIS 1800 and BV was preferred. The ARIS 3000 was used in low-frequency mode to collect data whose resolution is similar to the video recorded by the ARIS 1800.

As mentioned previously, the cross range resolution depends on the AC characteristics: the number of beams, the FOV opening and the range. It will therefore vary according to the types of AC. Figure 15 displays the beam resolution depending on the range for the three AC used in our study. Down-range resolution values, summarised in Table 1, are influenced by the number of samples. Their number may be chosen by the operators for the ARIS cameras but is fixed for the BV ones.

Pixel resolution is another characteristic that depends on the types of AC. Regarding the BlueView cameras, the size of the image is fixed. No matter the window length, the resulting AC images will have the same size. The pixel resolution will therefore be higher for smaller window lengths. Regarding the ARIS cameras, the down-range resolution can be set up by the operator and the pixel resolution might therefore vary.

Table 1 – Technical characteristics of the three AC used in this study. Their characteristics are displayed for both acquisition modes: the LF one and the HF one.

AC model	ARIS 1800		ARIS 3000		BlueView M900-2250 130	
Acquisition mode	HF	LF	HF	LF	HF	LF
Frequency	1.8 MHz	1.1 MHz	3 MHz	1.8 MHz	2.25 MHz	900 kHz
Frame rate (frames/sec)	3.5 to 15		4 to 15		5 to 8	
Number of beams	96		128		768	
Maximum range (m)	15	35	5	15	10	100
Minimum range (m)	0.7		0.7		0.2	
FOV opening (°)	28 x 14		28 x 15		130 x 20	130 x 12
Range resolution (mm)	3 to 23		3 to 19		6	13
Weight in air (kg)	5.82		5.12		2.5	
Dimensions (cm)	31 x 17 x 14		26 x 16 x 14		20.6 x 12.7 x 10.1	

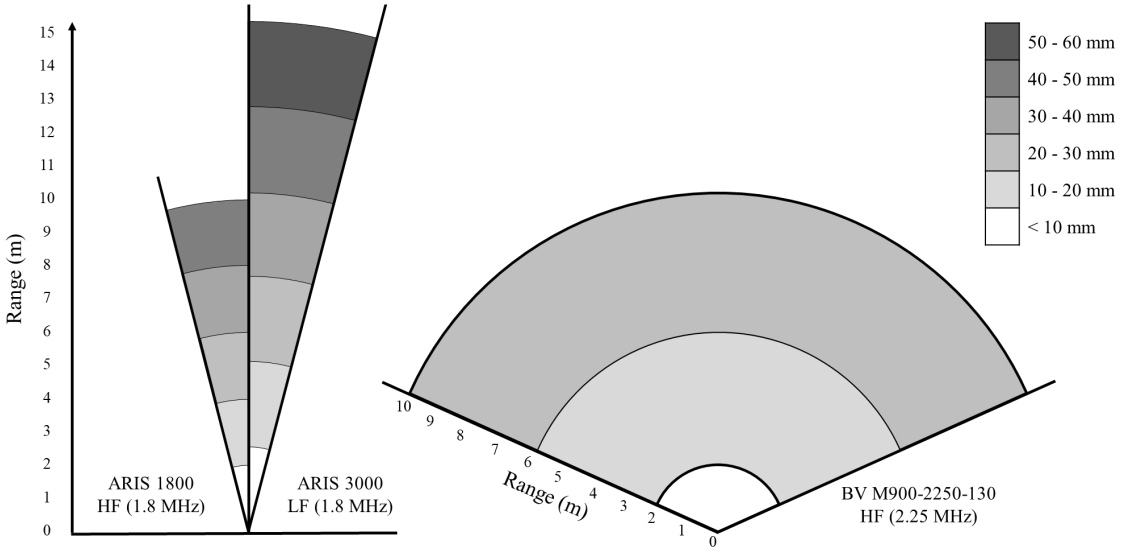


Figure 15 – Cross range resolution depending on the range for the ARIS 1800, ARIS 3000 and BV cameras used in this study. The cross range resolution has been calculated using (eq. 2).

The acquisition settings are therefore different from one AC to another (Table 2).

Table 2 – Acquisition settings for the ARIS and BlueView devices. In green, the parameters that can be chosen by the operator, in red, the ones that are fixed.

	ARIS	BlueView
Mode of acquisition (HF/LF)		
Opening of the FOV		
Range (Min/Max)		
Number of samples		
Down-range resolution		
Video volume		

2 Monitoring sites: Mauzac, Ducey and Port-La-Nouvelle

Data used in this study have been recorded at three monitoring sites located in France (Figure 16): the dam of Mauzac, the monitoring site of Ducey and the monitoring site of Port-La-Nouvelle.

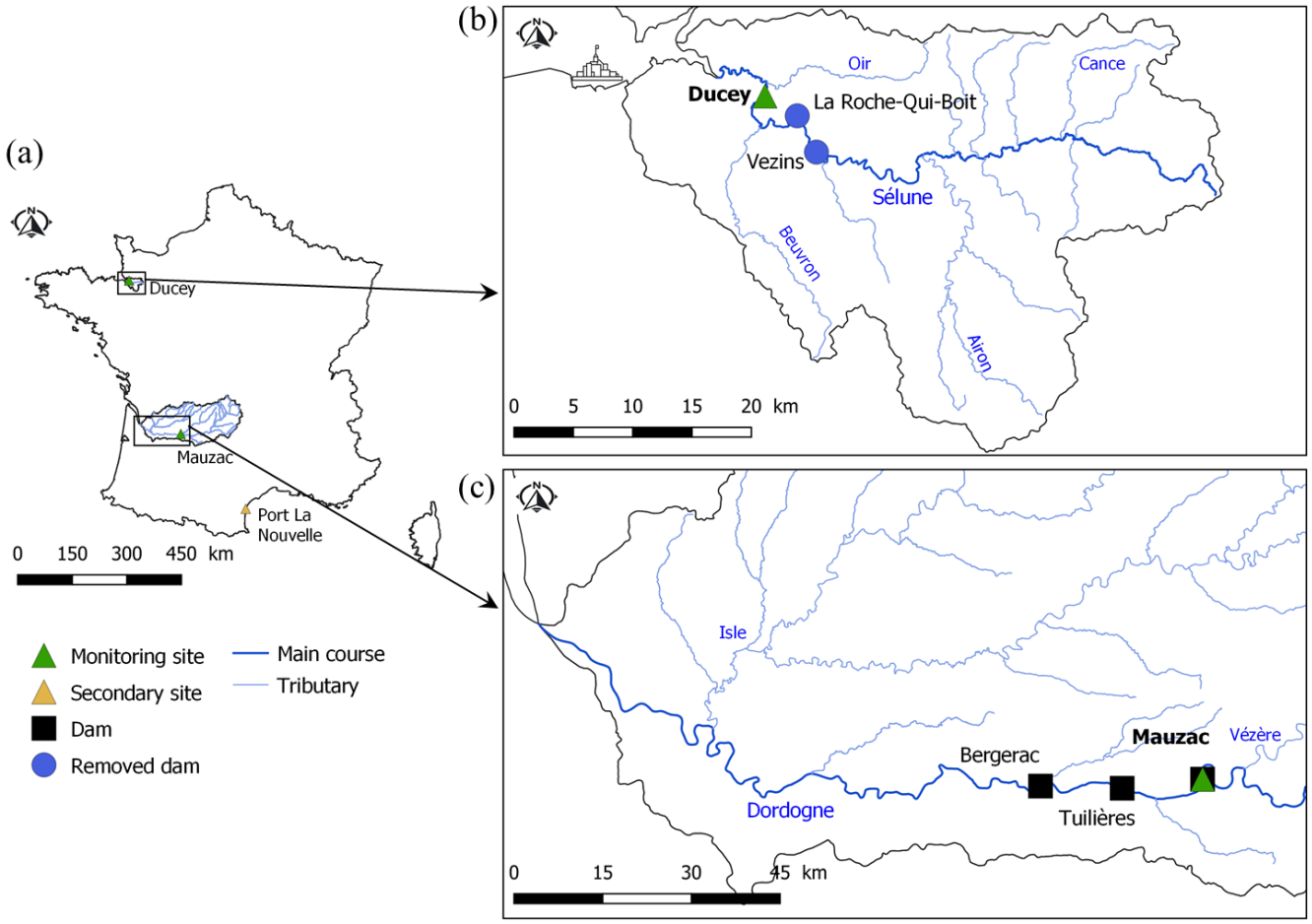


Figure 16 – Location of the two monitoring sites (a). The watershed of the Sélune River is displayed as well as the monitoring site of Ducey (green triangle) and the two former dams of La Roche-Qui-Boit and Vezins (blue circles) which were removed in 2018-2022 (b). The watershed of the Dordogne River is displayed as well as the monitoring site of Mauzac (green triangle) and the three dams of Bergerac, Tuilières and Mauzac (black squares) (c).

2.1 The Mauzac hydropower plant

The Mauzac hydropower plant study site is located in the South-West of France, on the Dordogne River (Figure 16.a). With a total length of 483 km, the Dordogne River has a watershed that covers 24 000 km^2. Located at 162 kilometres from the Gironde estuary (Figure 16.c), the dam of Mauzac is the first obstacle of the upstream part of the watershed that European eels encounter on their way to the sea. Mauzac facility is equipped with three fish ways: one on each riverbank of the dam (a vertical slot fish way and a one-baffle fish way) and a third fish way at the hydropower plant (a vertical slot fish way) (Figure 17). Mauzac inlet canal is almost one kilometre long and 50 meters wide and is located on the left bank upstream to the hydropower plant (Figure 17). The dam remains closed as long as the river flow is inferior to the hydropower plant design flow of 300 m^3/s. Above this value, the dam is progressively opened as the water flow increases, corresponding to the action of spillage.

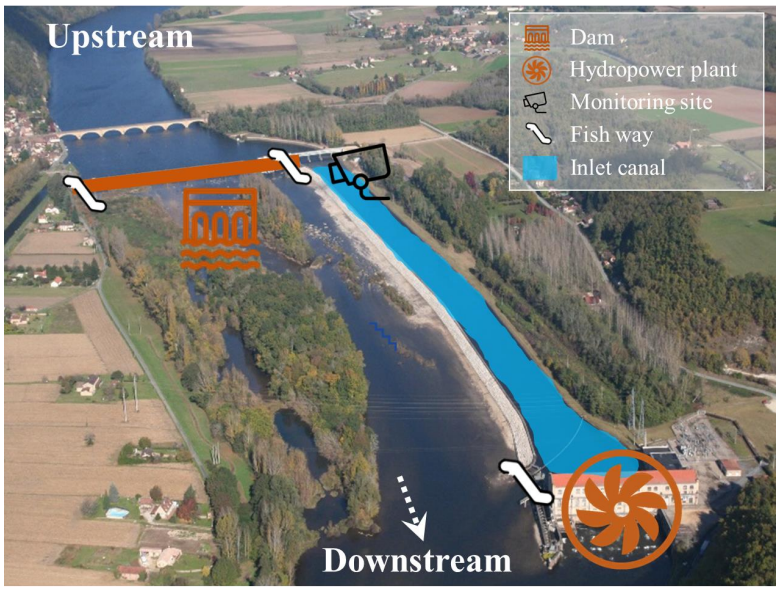


Figure 17 – Aerial view of the Mauzac dam, the dam and the hydropower plant are indicated in orange. The inlet canal is displayed in blue. The three fishways are displayed in white. The monitoring site where the AC have been installed is indicated in black, at the entry of the inlet canal.

Mauzac and the two consecutive downstream hydropower plants are exploited by EDF. They are carefully monitored during the seaward migration period of European eel to prevent turbine mortality. Radio telemetry campaigns have highlighted that with low water flow eels can pass through the inlet canal while they can pass through the dam with high water flow (Figure 18). Hence, above the hydropower plant design flow, the dam is open, and spillage is carried out, i.e., the gates of the dam are progressively opened as the river flow increases, becoming a favoured passage for eels [ECOGEA, 2014]. For eels passing through the inlet canal, scientific fisheries campaigns using nets attest the passages of eels on the left bank of the canal [ECOGEA, 2014]. However, radio telemetry campaigns also highlight that when the water flow increases eels shift towards the right bank of the inlet canal [ECOGEA, 2014].

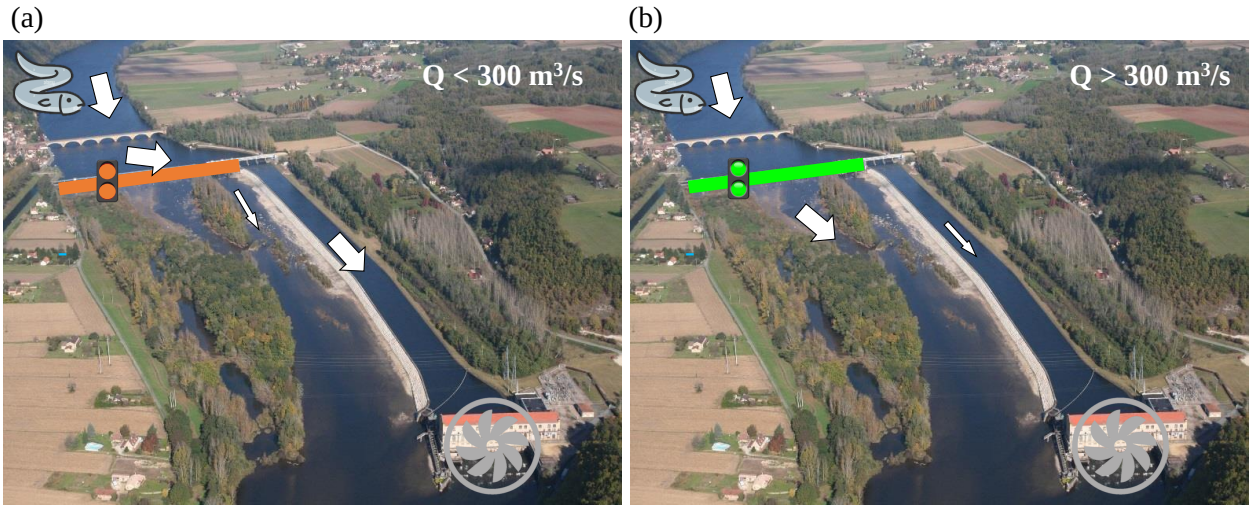


Figure 18 – Eels favoured passages at the Mauzac facility. The wider the arrow, the more the passage is favoured by eels. In case of water flow inferior to 300 m^3/s, the dam is closed and the eels pass through the inlet canal (a); however, if the water flow is superior to 300 m^3/s, spillage at the dam favours eels passages through natural river channel (b).

Eels captured are mainly large individuals of 60-90 cm long female ones (Figure 19). European eels being a species of high ecological interest, on the IUCN red list, turbine shutdowns are carried out at Mauzac as a measure of mitigation. Monitoring of their passages at this location is therefore of great interest for EDF in order to collect information on their migration and to optimise the mitigation measures applied. Other species are also present at this location, mainly Cyprinidae species: the bleak (*Alburnus alburnus*), the barbel (*Barbus barbus*), the common bream (*Abramis brama*), the silver carp (*Hypophtalmichtys molitrix*), the European chub (*Squalius cephalus*), the common roach (*Rutilus rutilus*) and the common dace (*Leuciscus leuciscus*) but also carnivorous species like the Northern pike (*Esox lucius*), the pikeperch (*Sander lucioperca*), the European perch (*Perca fluviatilis*) and the European catfish (*Silurus glanis*) [ECOGEA, 2014].

An ARIS 1800 camera has recorded continuously for three migration periods: from November 2014 to January 2015, from December 2015 to June 2016 and in December 2018. A BV camera has then been installed for the last monitored migration period (2018). Due to technical issues, the BV camera has been recording for a short period, in which four days have been also recorded by the ARIS. Both cameras have been installed close to each other (Figure 20), overlapping the water section they respectively cover. The ARIS 1800 camera was fixed to the dock, while the BV one was tied to a floating platform attached to the dock (Figure 21). Both cameras recorded data perpendicularly to the flow direction up to 10 meters for BV and to 9.4 meters for ARIS 1800. BV having a wider FOV opening than ARIS 1800, it covers a wider water volume. In the Mauzac configuration, BV covers 15 times the volume covered by the ARIS 1800 ($\approx 500m^3$ vs. $34m^3$) [De Oliveira and Le Quinio, 2022]. However, the two devices

cover a common proportion of the water section. Considering the first portion of the inlet canal width (15 meters), the ARIS 1800 and the BV respectively cover 32% and 29% of the water section (Figure 20).

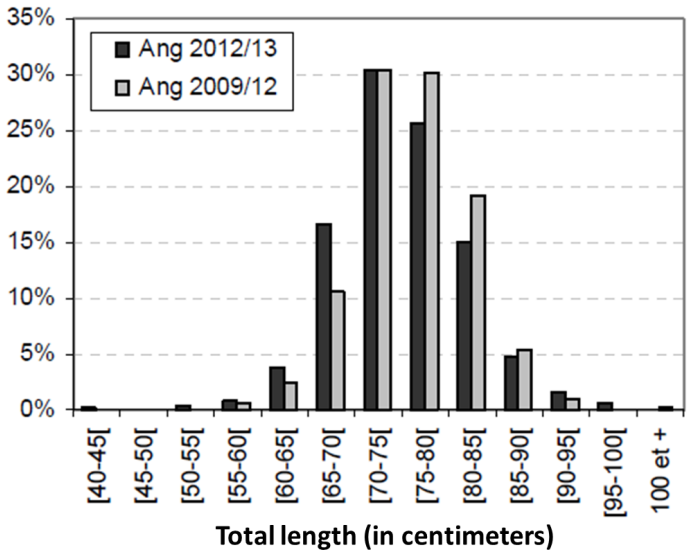


Figure 19 – Distribution (percentage) by size class (cm) of European eels caught between 2012 and 2013 (Ang 2012/13), between 2009 and 2012 (Ang 2009/12), at the Mauzac fishery, translated from French from [ECOGEA, 2014].

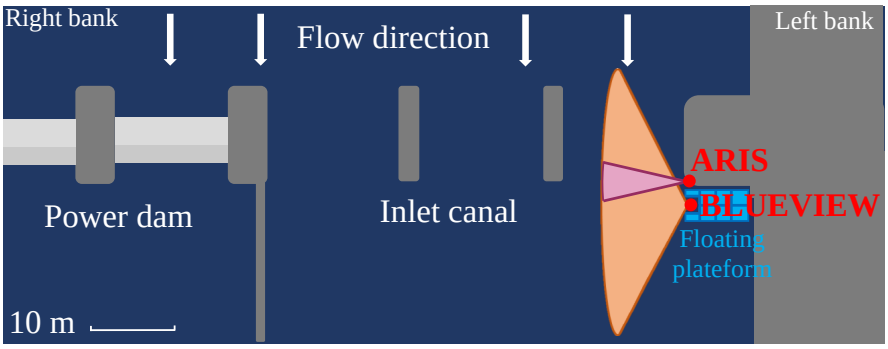


Figure 20 – Scheme of the installation of ARIS 1800 and BV cameras at the monitoring site of Mauzac. The dam of Mauzac is located on the right bank. The two AC have been installed on the left bank, at the entry of the inlet canal. The ARIS 1800 is fixed to the bank while the BV is fixed to a floating stationary platform. From [Le Quinio et al., 2023].

Figure 21 – Picture of the floating platform, on which the BlueView camera is fixed (left pink ellipse) and of the support of the ARIS 1800 (right pink ellipse).

2.2 The Ducey monitoring site

The Ducey study site is located in the North-West of France, on the Sélune River (Figure 16.a). With a total length of 85 km, the Sélune River has a watershed that covers 1038 km^2. Situated at the downstream part of the river (Figure 16.b) at about 10 km from the Baie du Mont Saint-Michel, Ducey is downstream the two former dams of La Roche-Qui-Boit and Vezins. These two dams have been recently removed (2018-2022) to restore the river continuum [Cholley et al., 2015]. AC are used here in Ducey to quantify migration dynamics of diadromous species before and after the removal of the dams [Martignac, 2016]. Many diadromous fish species can be observed at Ducey [Martignac, 2016], including the anadromous Atlantic salmon (*Salmo salar*), Sea trout (*Salmo trutta*), European River lamprey (*Lampetra fluviatilis*), Sea lamprey (*Petromyzon marinus*), Allis shad (*Alosa alosa*), twaite shad (*Alosa fallax*), and the catadromous European eel (*Anguilla anguilla*) and Thinlip Grey mullet (Chelon ramada). Other freshwater fish species are also observed: the European catfish (*Silurus glanis*), the Common carp (*Cyprinus carpio*), the Common bream (*Abramis brama*), the pikeperch (*Sander lucioperca*) and the Northern pike (*Esox lucius*). The continuous monitoring started in August 2013 [Martignac, 2016], allows the researchers to gather a consequent set of information on those species.

An ARIS 3000 has been set up for a few months in 2017 and 2019 (Figure 22). The camera has recorded continuously at low frequency mode (1.8 MHz). At the camera's location, the Sélune is 18 m wide and 1.8 m deep maximum. The ARIS 3000 unit has been tilted from 5° to 6° so that the beams do not hit the surface preventing videos from being disturbed by water surface echoes or floating objects such as leaves but it consequently hits the bottom of

the river. To cover a maximum of the water section, the detection beam of the ARIS 3000 was fixed, from 2 to 16.4 meters, corresponding to almost 85% of the water section.

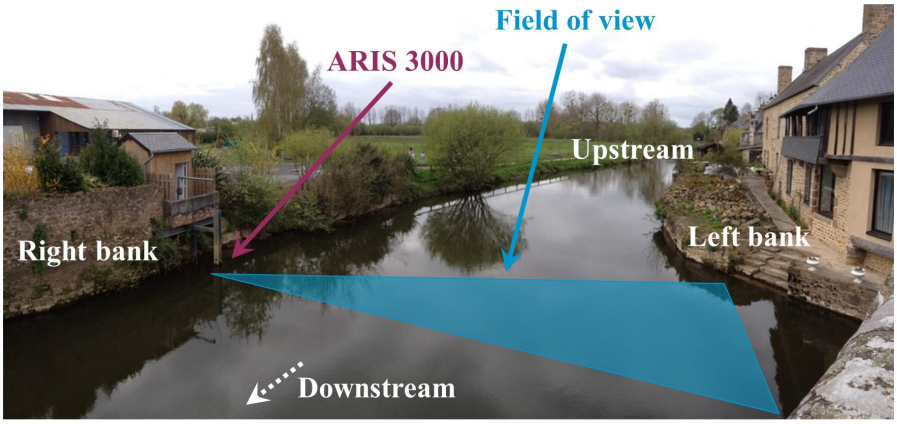


Figure 22 – Installation of the monitoring site of Ducey. The ARIS 3000 has been installed on the right bank covering up to 16.4 meters of the river 18-meters width. The FOV of the AC is represented as an approximation in the blue triangle.

2.3 The Port-La-Nouvelle monitoring site

Data recorded on a third monitoring site have been used in the paper published in PLOS ONE [Le Quinio et al., 2023]. This site is located at Port-La-Nouvelle (South-West of France, Figure 16.a) and is equipped with an ARIS 1800 camera, set to monitor European eels during their downstream migration. Monitoring objectives, site characteristics, and AC settings are described in details in [Lagarde et al., 2020] and [Lagarde et al., 2021].

Port-La-Nouvelle data are only used to assess the performances of the detection and identification methods developed in this PhD work and not to develop these methods. Consequently, this monitoring site is considered as a secondary site. The AC dataset and the database of eel countings were kindly provided by the CEFREM laboratory (UMR 5110, CNRS-UPVD, Perpignan, France), in charge of this monitoring. The contents of the dataset we used is described in [Le Quinio et al., 2023], in Chapter 3.

3 AC data visualisation: the operator truth

3.1 The operator truth

To design and evaluate a method of species identification, data labelled by experts is required. Passages of fish have therefore to be identified through direct observation, it is the ground truth. In the case of the AC data used in my PhD work, no ground truth is available: no direct observation of the fish is possible because they have only been recorded by the AC. To obtain a ground

truth, it would have been necessary to capture all of the fish passing in the AC FOV to identify their species and measure their length or to record preliminary identified and measured individuals during their passage in the AC beam [Daroux et al., 2019, Lagarde et al., 2020]. None of those experiments were easily feasible due to time and logistic constraints.

However, it is still possible to obtain a labelling of the AC data via operator visualisation. This labelling will be considered as the reference for the proposed work and is called the operator truth.

The labelling by the operator is made regarding its ability to distinguish fish species on the acoustic videos. Their discrimination relies on morphological and behavioural characteristics to identify the species. However, in some cases, both attributes are shared by several species, preventing any discrimination between them. Those groups of species are consequently described with a common label, not based on taxonomic considerations. In my work, data have been labelled by experienced operators, familiar with the acoustic videos and the morphology and behaviour of the species present in the monitoring sites. All of the species and groups of species considered for the visualisation are presented in the following subsection.

3.2 Groups of species for AC data visualisation

The EEL group is composed of European eels (Figure 25.a). European eel is a catadromous species that spawns in the Sargasso Sea [Tesch, 2007] (Figure 23). They reach the continental coast where they metamorphose to glass eels [van Ginneken and Maes, 2005] before migrating to freshwater rivers where they will grow, during 6 to 10 years, at the yellow stage [Tesch, 2007]. They are then metamorphosed to silver eels and start their downstream migration shortly after [Vøllestad et al., 1986]. This downstream migration takes place in the winter [Righton et al., 2016] but can extend from August to February and in some cases spring migrations are also observed. The downstream migration depends on the latitude where it takes place as well as on the hydrologic conditions [Stein et al., 2015]. Silver eels present a dimorphism in body length between males and females. The females of European silver eels are larger than males, from 50 to 100 cm, while male European eels usually measure between 35 and 46 cm [Durif et al., 2005]. European eels are easily recognizable by human eyes on acoustic videos due to their serpentine shape and their anguilliform swimming mode [Lenihan et al., 2019, Lagarde et al., 2020], as long as the frame quality allows it. Other diadromous species share those characteristics, among the ones present in the two monitoring sites, such as the European river lamprey and the Sea lamprey. However, their migration periods and direction (upward or downward) differ from the European eel [Kelly and King, 2001], allowing the operator to get a reliable identification. European catfish is another serpentine shape, however its large and flattened head, in addition to its length, for the larger individuals,

makes possible a discrimination with the European eel by the operator.

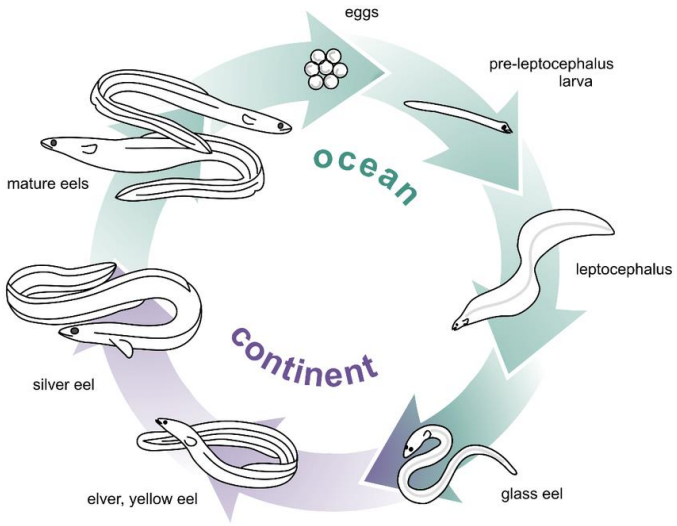


Figure 23 – Scheme of the European eel (*Anguilla anguilla*) life cycle from its spawns in the ocean to its maturation in freshwater environments before migrating back to its spawning location, from [Henkel et al., 2012].

The SAT group is composed of Atlantic salmon (*Salmo salar*) (Figure 25.b), a diadromous species whose life cycle requires migration between marine and freshwater habitats (Figure 24). Atlantic salmon is an anadromous fish that breeds in freshwater environments. After one to three years in rivers, young salmons metamorphose to become smolts and start, in spring, their downstream migration from foraging areas to the sea. After one to four years at sea [Hutchings and Jones, 1998], adult salmons begin their upstream migration along the freshwater rivers back to spawning areas. The Atlantic salmons composing this SAT group are adults, recorded during their upstream migration. They measure between 50 and 80 cm and have a subcarangiform swimming mode. Among the adult salmon, the grilse, usually ranging from 55 to 70 cm, spends one single winter at sea before returning to fresh water from June to October [Mills, 1971, Jonsson et al., 1990, Amundsen et al., 2003]. Other individuals spend more than one winter at sea and are thus larger, from 65 to 90 cm, while returning to freshwater earlier in the season, from March or April [Mills, 1971, Jonsson et al., 1990, Baglinière and Porcher, 1994, Amundsen et al., 2003]. The Atlantic salmon might be confused with the Sea trout, another anadromous species, that has similar behavioural and morphological characteristics, their length distributions overlapping. However, on the Sélune river watershed, more than 90% of Sea trout are shorter than 50 cm [Martignac, 2016].

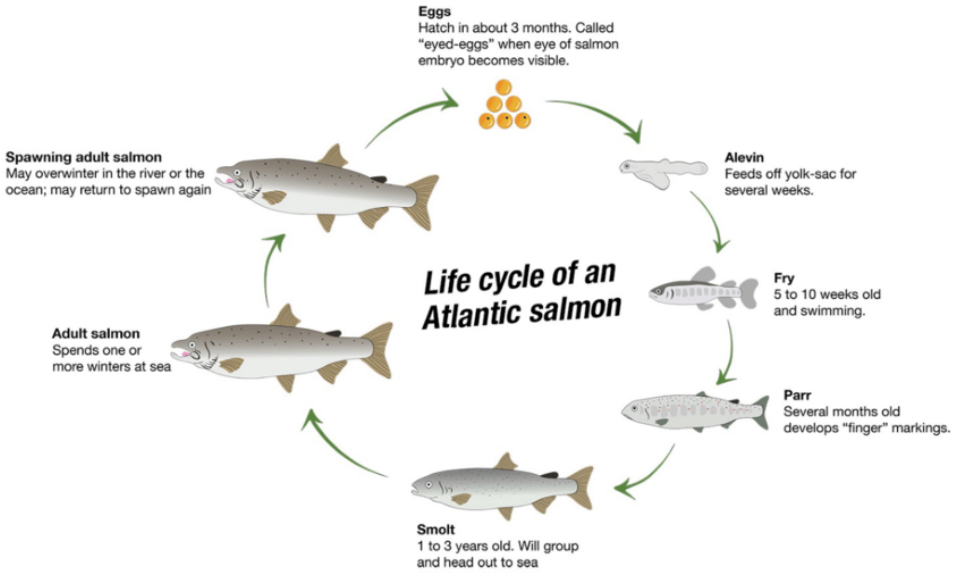


Figure 24 – Scheme of the Atlantic salmon (*Salmo salar*) life cycle from its spawns in freshwater rivers to its maturation in the ocean before migrating back to its spawning location, from [ParksCanada, 2022].

The SIL group corresponds to the European catfish (*Silurus glanis*) (Figure 25.c). The European catfish is a resident fish that colonises freshwater rivers. This species is an opportunistic predator whose presence raised management concerns [Cucherousset et al., 2018], especially regarding anadromous migratory fish that may represent an important part of their diet [Syväranta et al., 2009], among them the Atlantic salmon [Boulêtreau et al., 2018]. It has a negative impact on the migrating fish nearby the fish passage structures that act like funnels. The adult European catfish is particularly recognizable because of its large size, it can reach up to 2.5 m [Keith et al., 2021]. The European catfish has an anguilliform swimming mode and its shape with a triangular head makes him recognizable on AC data [Boulêtreau et al., 2018].

The CYP group gathers common bream (*Abramis brama*) and common carp (*Cyprinus carpio*), both from the Cyprinidae family (Figure 25.d). Both species are potadromous fish, staying in freshwater environments their whole life. The adults can reach similar length, between 28 and 50 cm for the common bream, and between 50 and 80 cm for the common carp [Keith et al., 2021]. Besides, they have a similar foraging behaviour that allows the operator to distinguish them from other species.

The PRED group is composed of pikeperch (*Sander lucioperca*) and Northern pike (*Esox Lucius*) (Figure 25.e). Similar to the CYP, PRED are potadromous fish. The adults reach almost the same range of length than CYP, mainly between 50 to 60 cm for the adult Northern pikes, and 35 and 40 cm for the male and female adult pikeperch respectively, but both may measure up to 90-100 cm [Keith et al., 2021]. However, their morphology differs from CYP, PRED hav-

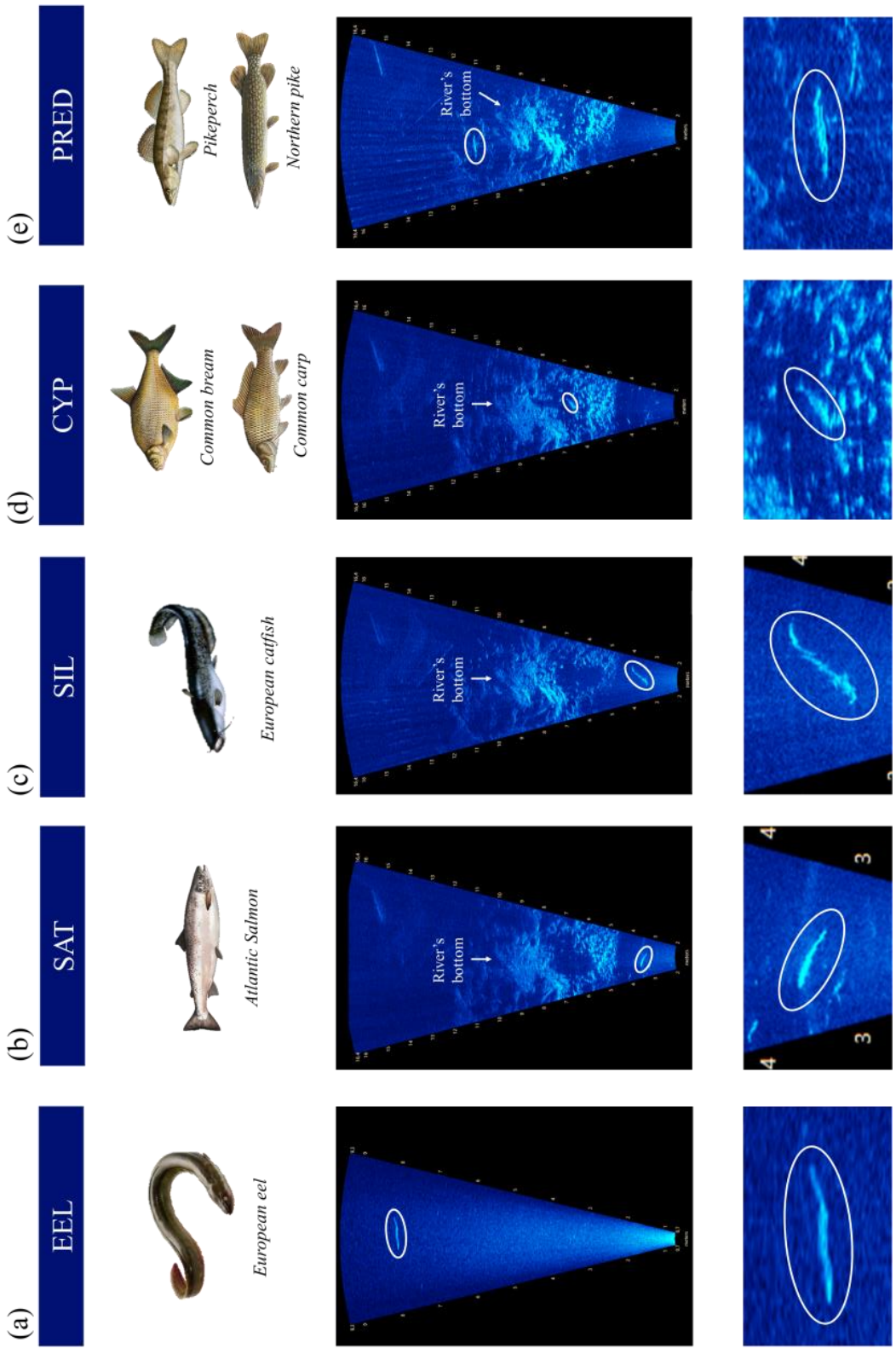


Figure 25 – Examples of the images captured by AC. The columns correspond respectively to the EEL (a), SAT (b), SIL (c), CYP (d) and PRED (e) groups of species respectively. The first layer displays pictures of the species, the second layer has been recorded at Mauzac using an ARIS camera while the other images have all been recorded on the Sélune river using an ARIS camera.

ing a thinner body shape. Both have also recognizable dorsal fins, at the back for the Northern pike, and two dorsal fins for the pikeperch. Their behavioural characteristics are also different as they are predatory fish, hunting from a hide and feeding from small fish. The Northern pike has however a very distinctive hunting behaviour as it is able to remain stationary in the water until its prey appears. In case this hunting behaviour is recorded on videos, distinction with pikeperch is possible. Otherwise, acoustic videos do not often provide enough details on the morphology of fish, such as the fins' presence and positions, to allow the operator to make the discrimination. The PRED group has therefore been defined to integrate this discrimination difficulty on acoustic videos, despite the pikeperch and Northern pike not belonging to the same taxonomic family.

3.3 Labelled data for species identification

3.3.1 From acoustic videos to the datasets

Despite pursuing the same aim of labelling data, the protocol of visualisation varies between monitoring sites. AC data of the two main monitoring sites of this PhD work have been partially visualised by experienced operators:

- **Mauzac monitoring site:** the operators have used VLC Media Player to visualise acoustic videos. For each eel passage detected, its date and hour, as well as the numbers of the first and last frame in which the eel can be seen, were listed.
- **Sélune monitoring site:** the operators have used the ARIS Fish software 2.6.3 that allows them to retrieve additional information on fish passages. For each of them, the date and hour of the passage, the first and last frame of detection have been noted. The fish length has been measured with the ARIS Fish software on the frame considered as the most representative of the fish length observed during its detection by the operator. The average distance from the sonar (or range) and the direction of the fish have also been retrieved. When it can be formally identified by the operator, the species is annotated according to a decision tree based on morphological and behavioural characteristics of the fish during its detection (Figure 26). Additional comments may complete the information collected for each fish passage.

During their data reviewing, operators regularly observed some recurrent artefacts that may hide fish passages. Arcing effects appear on the ARIS 1800 and ARIS 3000 frames. They are artefacts of high intensity resulting from returning echoes, which deform fish by giving them a more streamlined shape [Mueller et al., 2008] or hiding their echoes (Figure 27.a). BV videos are also subject to artefacts as a continuous structural noise of unknown source is observed at every 21° of the FOV (Figure 27.b).

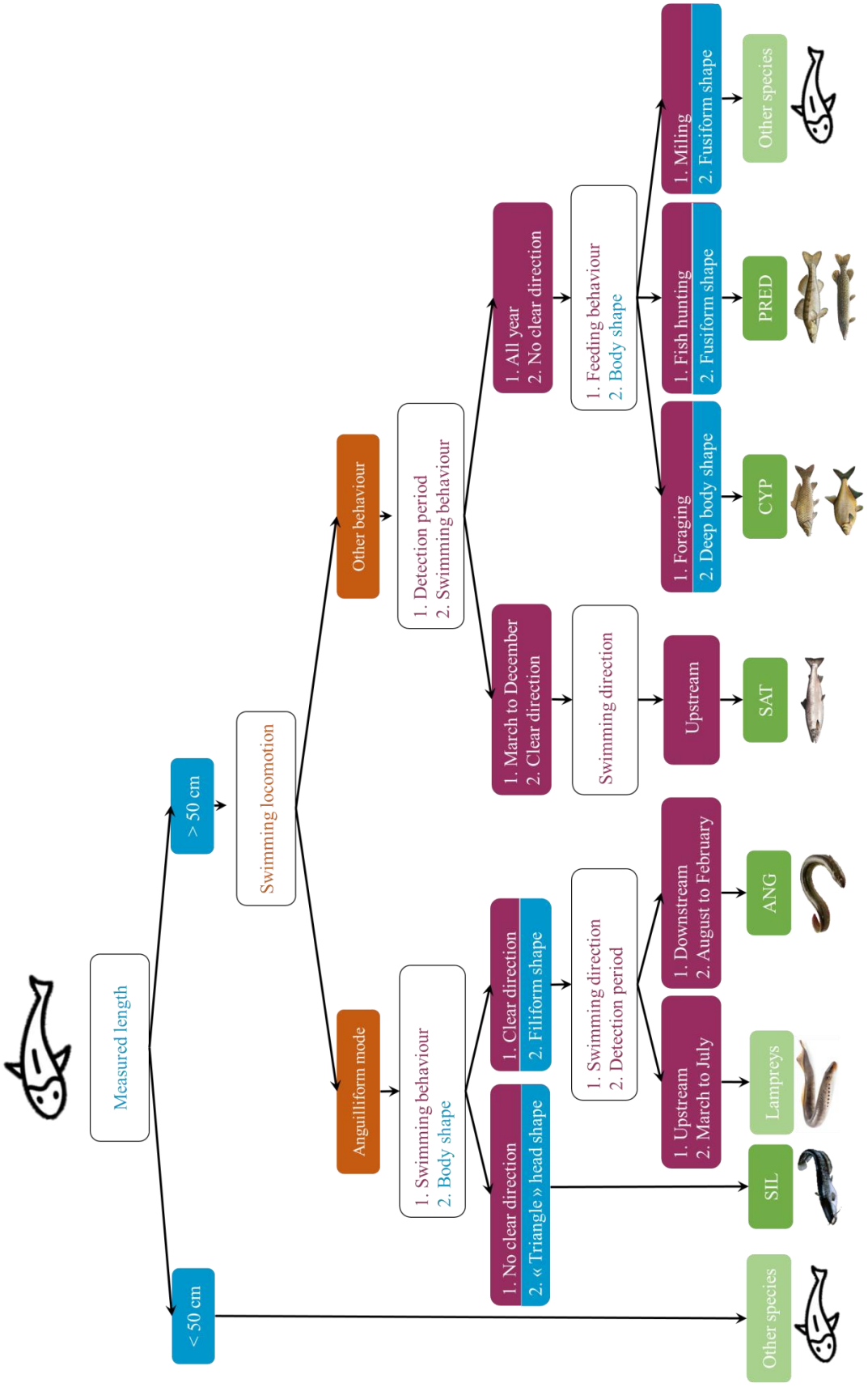
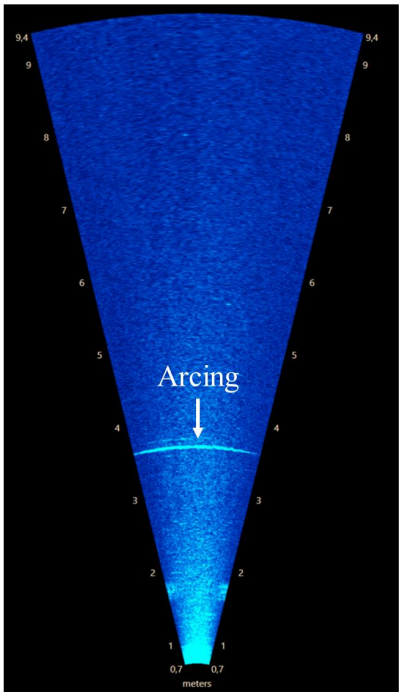


Figure 26 – Decision tree followed by an operator to identify fish species thanks to its morphological (in blue), swimming locomotion (in orange) and behavioural (in purple) characteristics.

(a)



(b)

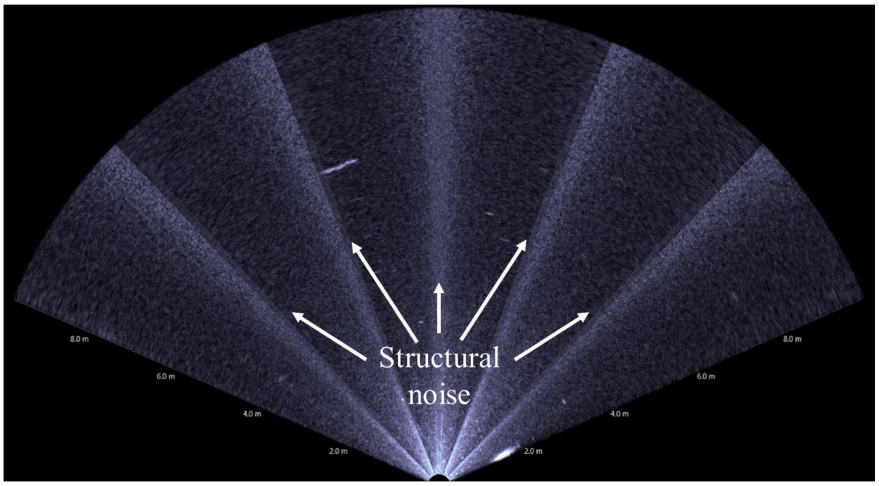


Figure 27 – A screenshot of a video recorded at Mauzac using an ARIS camera showing arcing effect (a), and, a screenshot of a video recorded at Mauzac using a BV camera highlighting the structural noise observed.

3.3.2 Datasets

Three datasets have been built from the acoustic videos recorded at the two main monitoring sites of this PhD work, with the two types of AC: MZC-ARIS (Mauzac, ARIS 1800), MZC-BV (Mauzac, BlueView) and SEL-ARIS (Sélune, ARIS 3000). Each dataset is composed of the operator reviewed data. Both MZC-ARIS and MZC-BV are eel-specific datasets as the Mauzac monitoring focus on this species, while the SEL-ARIS dataset lists the passages of several species or groups of species. All of the characteristics of the datasets have been summarised in Table 3.

MZC-ARIS videos have been recorded in November 2014 and in December 2018 when peaks of eels' downstream migration have been observed leading to a high amount of recorded eels passages on the acoustic videos. MZC-BV have been recorded in December 2018 with four days in common with MZC-ARIS. Regarding the SEL-ARIS multi-group dataset, videos correspond to the recording period of May to September 2019. SEL-ARIS has an unbalanced distribution of the group's sizes (Table 3) due to higher difficulty for the operator to distinguish some of the groups from the others, in particular for the CYP group; or by their less frequent detection in the acoustic videos which is the case of the PRED group.

Table 3 – Description of the three datasets MZC-ARIS, MZC-BV and SEL-ARIS regarding their AC settings and the characteristics of the monitoring site and studied species or groups of species.

	MZC - ARIS	MZC - BV	SEL - ARIS
Monitoring site	Mauzac	Mauzac	Ducey
AC models	ARIS 1800	BlueView M900-2250 130	ARIS 3000
Frequency (MHz)	1.8	2.25	1.8
Number of beams	96	768	128
Number of samples	384	1630*	512
Beam resolution (mm)	5 to 50	3 to 30	4 to 60
Sample resolution (mm)	22.6	6	28
Range (m)	0.7 to 9.4	0.2 to 10	2 to 16.4
Frame rate (frames/sec)	7	5	5
Frame height (px)	1276	1238	1350
Frame width (px)	664	2302	738
Pixel resolution (mm)	7	11	8
Operator analysed period of time	November 2014 & December 2018	December 2018	May to September 2019
Number of ANG	783	198	0
Number of SAT			39
Number of CYP			6
Number of SIL			34
Number of PRED			2

These datasets are used to design and evaluate the pipeline of acoustic imaging automatic analysis proposed in this manuscript.

Chapter 2. Sorting data of interest: Detection and tracking, the first step to species identification

Contents

La science, mon garçon, est faite d'erreurs,
mais d'erreurs qu'il est bon de commettre,
car elles mènent peu à peu à la vérité.
Jules Verne

1 Introduction

This chapter focuses on sorting the data of interest from the high amount of data recorded by the AC through the proposed pipeline. In case of fish migration studies, our data of interest are targets passing through camera FOV that might be fish. The target is defined as an area of contiguous pixels of similar intensities that can be tracked from one frame to another, along its trajectory in the camera FOV. The three-steps process guideline (Figure 28) is the following:

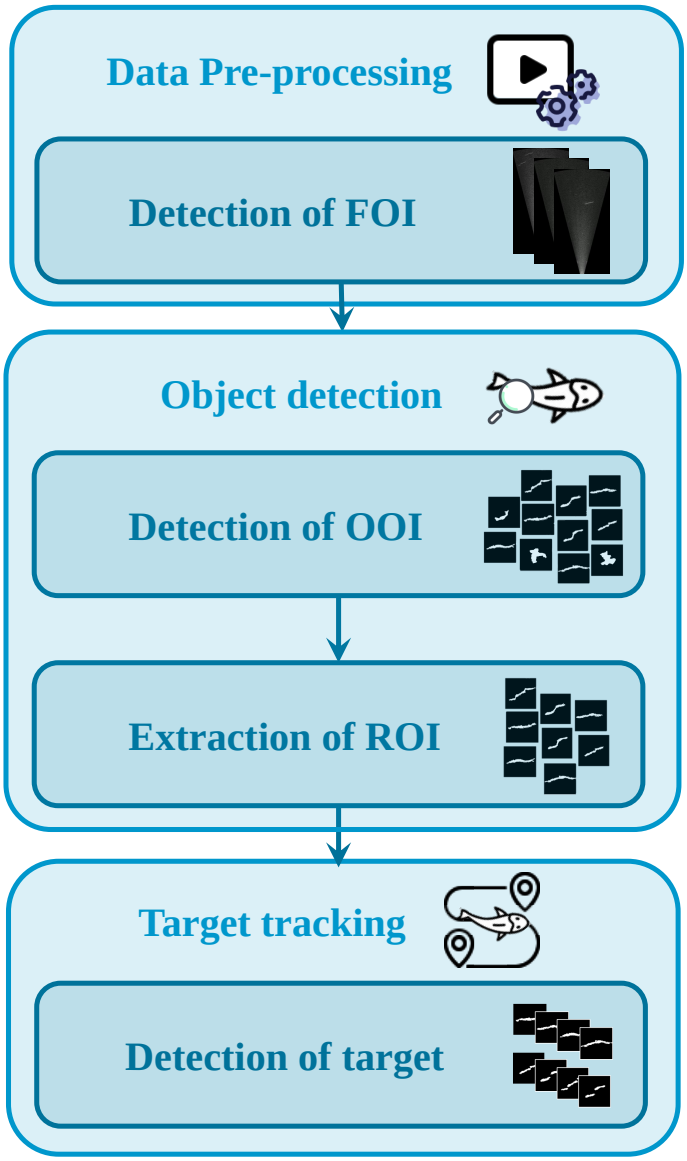


Figure 28 – Proposed pipeline, from the extraction of FOI (Frames Of Interest), the detection of OOI (Objects Of Interest), the extraction of ROI (Regions Of Interest) to the detection of the targets.

- **Extraction of the Frames Of Interest (FOI) sequences.**
  The first step is to discriminate the FOI in which the Objects Of Interest (OOI) are passing. At this step of the process, Objects Of Interest (OOI) correspond to regions of contiguous pixels of similar intensities from the frame's foreground, moving from one frame to another.

- **Extraction of the Regions Of Interest (ROI).**
  All of the OOI do not correspond to a fish passage. The second step of our process consequently aims to filter the OOI present on the FOI. Furthermore, to be efficiently tracked, the selected OOI are post-processed to get a representation of them as a satisfying single continuous region of pixels. The resulting images correspond to the ROI.

- **Extraction of the passages of the targets.**
  The ROI are then tracked from one frame to the other along their entire trajectory in the camera FOV. Each complete tracking becomes one target that might correspond to a fish. The target is recorded and detected on successive images that correspond to the successive ROI tracked.

In this chapter, an automatic image processing pipeline is introduced. It has been developed to complete each of these three steps. The choices made to design the pipeline and the improvement it could bring to acoustic videos analysis are also discussed.

2 The aim of a generic pipeline

2.1 Conversion of acoustic videos to a generic video format

Each type of AC records data into its own format that cannot be handled with programming languages such as Python. Prior to the first step of our process and to ensure the genericity of the analysis of the different types of AC, a conversion from manufacturer video format to a generic video format, the AVI, must be done. Python codes have been developed by EDF R&D to automatically convert ARIS and BlueView videos to grayscale AVI videos and are available online from GitHub source: https://github.com/fzao/sonaris. The conversion code designed for the BlueView requires a computer package called *bvtsdk* purchasable from Teledyne Inc. It handles the extraction of all of the frames composing an acoustic video. The extraction of ARIS frames can be directly assessed with the developed Python code.

This conversion from the manufacturer format to AVI videos has been added to our pipeline as a preliminary step.

2.2 Generic parameters for detection and identification

Parameters are used in the proposed pipeline to carry out the detection, the tracking (Chapter 2) and the identification (Chapter 3 and 4) of the targets. Those parameters, expressed in millimetres or square millimetres, correspond to proper characteristics of the target such as its length (expressed in millimetres) or surface area (expressed in square millimetres). The aim is to get rid of the pixel resolution that relies on the AC acquisition settings and that may vary from one monitoring site to another or from one type of AC to another. A conversion from pixel to millimetres is thus done thanks to the pixel resolution (eq. 3).

To improve the detection of the target, a variable L_{Target} is arbitrary defined according to the morphological characteristics of the species the operator is looking for. This value is chosen according to the target average fish length and enables a large interval of length around this value.

3 From entire acoustic videos to frames of interest

3.1 Calculation of AC echogram: an overview of the literature

The idea of examining FOI has been first addressed by the software ARIS Fish (Sound Metrics Corp., Bellevue, WA, USA) that proposes an echogram calculation from the acoutic videos. Its visual reading allows the operator to identify fish passages along a representation of a whole AC video into a single image [Mueller et al., 2008]. Regarding the AC echogram of the ARIS Fish Software, each frame is summarised as a matrix of a single column for which each row corresponds to the maximum of intensity observed in the corresponding row of the frame. The echogram concatenates each of these matrices to summarise the entire video into a single image (Figure 29). The horizontal axis can therefore be considered as the timeline while the vertical axis corresponds to the window length of the AC. No automatic analysis of the echogram is made yet by the software. However, the reading of such an echogram might be a key element to extract the FOI of the videos and remove from the analysis the other frames which will not provide data of interest for the following processing steps.

We therefore aim to develop a method combining echograms calculation with their automatic reading to retrieve FOI.

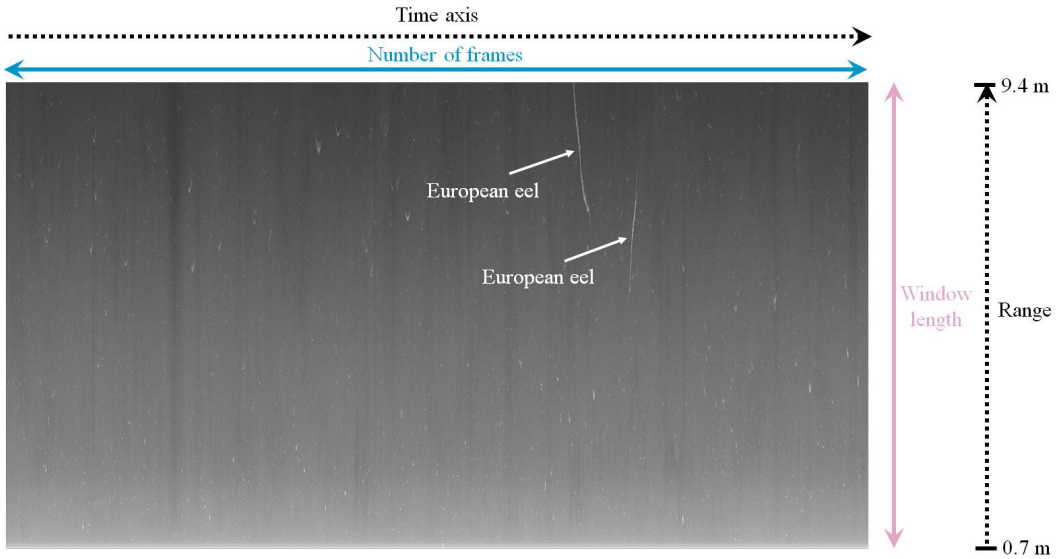


Figure 29 – Example of an echogram produced by the ARIS Fish software for a video from the MZC-ARIS dataset with two passages of European eels highlighted according to the operator truth. The vertical axis corresponds to the range axis while the horizontal axis corresponds to the time axis. The height and width of the echogram correspond to the window length and the number of frames of the video respectively.

3.2 Focus on anguilliform fish

To extract the FOI of a video, the intervals of frames during which an OOI moves in the camera FOV must be identified. Each OOI passage creates a discontinuity that corresponds to a different pixel intensity from the background pixels. The developed method synthesises the entire video in the format of an echogram in which only discontinuities are present.

As explained in previous chapters, we first focus on the detection of FOI of anguilliform species passages at the Mauzac monitoring site (MZC-ARIS and MZC-BV).

3.2.1 Proposed method

The acoustic frames being noisy images, with salt and pepper noise, smoothing them allows us to attenuate that noise and favour the detection. A smoothing of the frame is therefore first carried out using a median filter (filter size $k = 50$ mm) (eq. 4) [Ahmad and Sundararajan, 1987], known to remove salt and pepper noise while preserving sharp edges [Justusson, 1981]. Each pixel is assigned the median of its neighbourhood included in a square window of size k x k.

$$\forall\, u \in [1, D_1],\ \forall\, v \in [1, D_2],\ I'(u,v) = median(I(u-k:u+k, v-k:v+k)) \qquad \text{(eq. 4)}$$

With I the input frame; I', the smoothed frame; k, the size of the median filter. We choose k so that the product of k and the pixel resolution r is equal to 100 mm.

Each frame is then analysed by performing a succession of Singular Value Decomposition (SVD) [Golub and Reinsch, 1969]. Square sliding windows scan the frame with 50% overlap (Figure 30). Window size, d_w, is defined according to the frame resolution r and to the minimum length targeted L_{Target} (eq. 5). Their number, W_1 along the horizontal axis and W_2 along the vertical one, thus depend on the frame dimension D_1 x D_2 (eq. 6).

$$d_w = \frac{L_{Target}}{r} \quad \text{(eq. 5)}$$

$$W_1 = 2 \cdot \frac{(D_1 - d_w)}{d_w}, W_2 = 2 \cdot \frac{(D_2 - d_w)}{d_w} \quad \text{(eq. 6)}$$

With L_{Target}, the targeted length of the fish to detect and given as an input by the user, in millimetres; r, the pixel resolution, in millimetres; D_1 and D_2, the size of the frame, in pixels; d_w, the square sliding window side dimension, in pixels; W_1 and W_2 the number of windows along the horizontal and vertical axes.

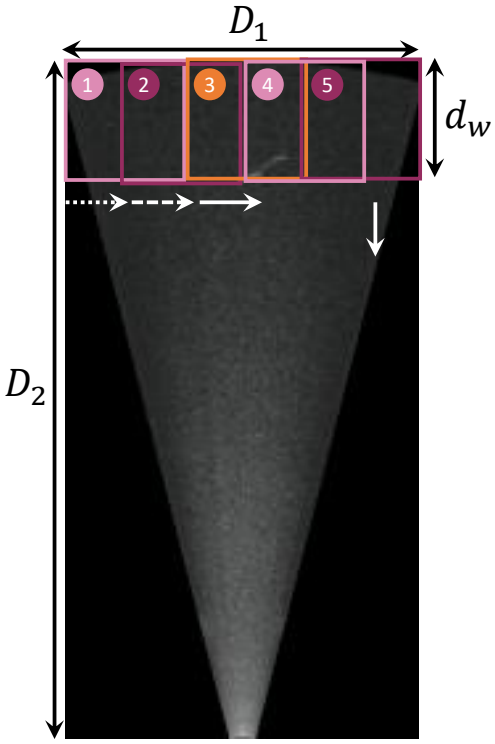


Figure 30 – Example of an acoustic frame after application of the median filter [Ahmad and Sundararajan, 1987], with squared sliding windows process highlights in colour. The white arrows show the direction of the window for the first slides. The width of the windows d_w as well as the size of the frame D_1 and D_2 are also indicated on the frame.

A SVD is thus systematically carried out on the successive parts of the image covered by the sliding windows. At each slide of the window, the ratio of each singular value (Singular Value Ratio, SVR) is calculated according to a reference image (eq. 7). The N^{th} SVR of a window

corresponds to the difference between (a) the ratio of the Nth Singular Value (SV) of the studied image over the sum of the N-1 SV, and (b) the ratio of the Nth SV calculated in the same way but on the reference image. This reference image is a standard image obtained from the same AC with no objects passing through, thus without any discontinuities. Successive SVR are recorded in a matrix of dimension W_1 x W_2, with one matrix for one SVR (Figure 31.a and Figure 31.c). To reduce the impact of the camera's edge effects, a mask is calculated relatively to the reference image (eq. 8). It is then applied (eq. 9) on each SVR matrix.

$$\forall w \in [1, W_1], \forall v \in [1, W_2], S_{w,v} = \frac{\sigma_N^{w,v}}{\sum_{n=1}^{n=N} \sigma_n^{w,v}} - \frac{\sigma_{N,ref}^{w,v}}{\sum_{n=1}^{n=N} \sigma_{n,ref}^{w,v}} \quad \text{(eq. 7)}$$

$$\forall w \in [1, W_1], \forall v \in [1, W_2], M_{w,v} = min(I_{i,j,ref})_{\substack{(w-1)*\frac{d_w}{2} \leq i \leq (w-1)*\frac{d_w}{2}+d_w \\ (v-1)*\frac{d_w}{2} \leq j \leq (v-1)*\frac{d_w}{2}+d_w}} \quad \text{(eq. 8)}$$

$$\forall w \in [1, W_1], \forall v \in [1, W_2], R_{w,v} = S_{w,v} \cdot M_{w,v} \quad \text{(eq. 9)}$$

With W_1 and W_2 the number of windows along the vertical and horizontal axes; S, the ratio matrix; $\sigma_N^{w,v}$, the N^{th} singular value of the window (w,v); $\sigma_{N,ref}^{w,v}$, the N^{th} singular value of the window (w,v) of reference; M, the mask; I_{ref}, the reference image; d_w, the square sliding window side dimension, in pixels; R, the ratio result matrix.

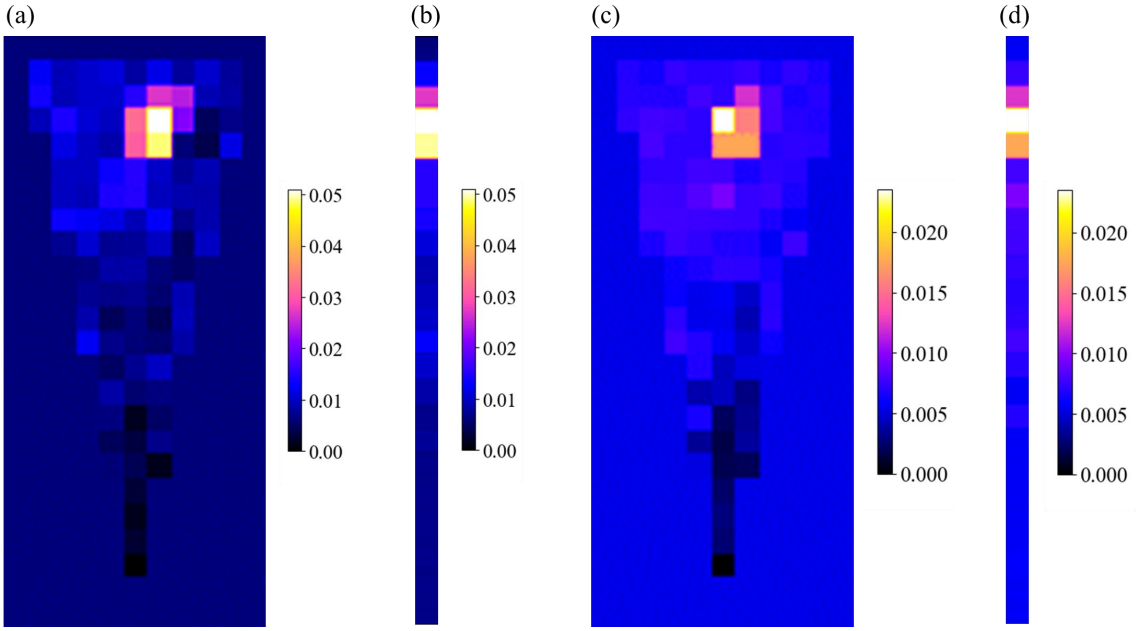


Figure 31 – Examples of the W_1 x W_2 matrices generated after frame process by the sliding windows and SVD calculation. One square corresponds to the ratio of SVR from one window. The two examples correspond to the ratio of the 2^{nd} SVR (a) and to the ratio of the 3^{rd} SVR (c). Examples of two matrices of size W_1x1 corresponding to the maximum ratios of the 2^{nd} SVR (b) and the 3^{rd} SVR (d) calculated along the horizontal axis.

From these resulting matrices, the maxima of the discontinuities ratios (eq. 10) along the vertical axis are calculated (Figure 31.b and Figure 31.d), i.e., along the main axis of the camera FOV. This allows us to study the discontinuities on the entire window length of the camera, perpendicularly to the axis of the flow. These maximum values progressively build up the echogram for each SVR studied (Figure 32). One column thus represents one frame of the video. Echogram's dimension matches the number of frames of the video by the number W_2 of sliding windows along the vertical axis of the frames.

$$\forall i \in [1, W_2], V_{i,1} = max(R_{i,j})_{1 \leq j \leq W_1} \quad \text{(eq. 10)}$$

With W_1 and W_2, the number of windows along the vertical and horizontal axes; V, the matrix of maximum ratios along the vertical axis; and, R, the ratio result matrix.

Only the three first SV are analysed as their corresponding egeinvectors explained more than 98% of the information of the sliding windows of the videos of test. It therefore corresponds to the analysis of the 2^{nd} and the 3^{rd} SVR according to eq. 7. Both echograms are processed by thresholding, using the threshold values for the 2^{nd} and the 3^{rd} SVR echograms (α_{SVR2} and α_{SVR3} respectively). Values of α_{SVR2} and α_{SVR3} therefore influence the resulting frame intervals and the performance of this step of the process. Different values of α_{SVR2} and α_{SVR3} will be tested to maximise the number of FOI with fish passages while minimising the total number of FOI.

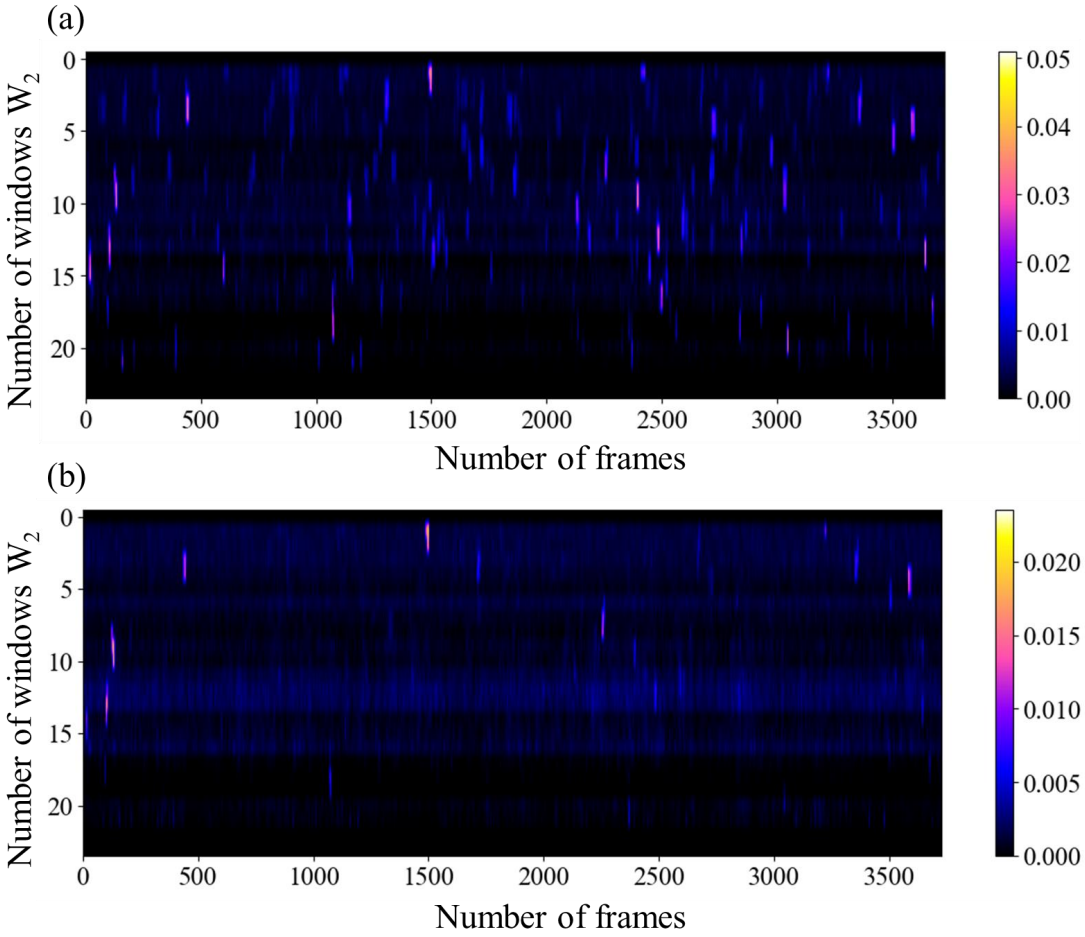


Figure 32 – (a) Echogram generated by the 2^{nd} SVR, (b) Echogram generated by the 3^{rd} SVR. The horizontal axis corresponds to the time axis, one column representing one frame. The vertical axis corresponds to the number of windows used to calculate the echogram, along this axis. The gradient of colour displays the value of the SVR, the lighter it is displayed the higher the value is.

3.2.2 Metrics of performance

The performance of the method is first assessed through the visual comparison between the operator truth N_{Fish} and the number of fish passages retrieved on the echogram before extracting the FOI.

To evaluate the method performance through the calculation of recall and efficiency values, 50 eel passages (25 from MZC-ARIS dataset and 25 from MZC-BV dataset) are used. The recall of the FOI detection (eq. 11) evaluates how accurate the detection of the frames, containing the fish we are looking for, is. The efficiency of the FOI detection (eq. 12) evaluates the number of FOI detected and therefore how filtering the method is. The aim of the method is to extract all of the frames in which fish of interest are present while minimising the number of FOI detected.

$$Recall = \frac{N_{DetectedTarget}}{N_{Fish}} \quad \text{(eq. 11)}$$

$$Efficiency = 1 - \frac{N_{FOI}}{N_{Frame}} \quad \text{(eq. 12)}$$

With $N_{DetectedTarget}$, the number of targets passages detected by the method; N_{Fish}, the total number of fish passages based on the operator truth; N_{FOI}, the number of FOI detected by the method; N_{Frame}, the total number of frames contained in the videos of the dataset.

As explained previously, the reading of our anguilliform-specific echograms relies on two complementary parameters, α_{SVR2} and α_{SVR3} that both influence the performance of the method. α_{SVR3} will be firstly tuned to maximise the recall. Once the value of α_{SVR3} fixed, α_{SVR2} will then be tuned to maximise both recall and efficiency.

The computation time required to generate echograms based on the proposed method is also indicated as it is a complementary way to measure performance.

3.2.3 Performance and tuning of the anguilliform-specific method

Figure 33 displays the examples of echograms obtained using the anguilliform-specific method on MZC-ARIS and MZC-BV datasets respectively. Red squares identify the passages of interest we are looking for, based on the operator truth. It highlights that the echograms of the 2^{nd} SVR displays more discontinuities than the echograms of the 3^{rd} SVR. However, for both SVR, the fish passages of the operator truth are visually present in the echograms.

Tuning of both thresholds α_{SVR2} and α_{SVR3} are then carried out to evaluate the FOI detection performance of the method. Figure 34 displays the recall reached when analysing only the 3^{rd} SVR echograms according to the value of α_{SVR3}. Performances differ between datasets. To optimise the performance, we thus choose α_{SVR3}=0.5 for the MZC-ARIS dataset and α_{SVR3}=1 for the MZC-BV dataset, the recall curves decreasing above this threshold (Figure 34).

Once the value of α_{SVR3} is known, the value of the second threshold α_{SVR2} can be tuned. The automatic reading of the echogram is performed. The number of fish passages detected, as well as the number of FOI, are retrieved for each value of α_{SVR2} tested. As for the optimal value of α_{SVR3}, the tuning of α_{SVR2} seems to be different on the two datasets. Regarding the MZC-ARIS dataset, Figure 35.a highlights a plateau of the recall around 92% until α_{SVR2} equal to 2, before a quick drop of the value. Figure 35.b highlights a quick increase of the efficiency from 0% to 75% until α_{SVR2} is equal to 1.5. The increase continues less quickly for larger α_{SVR2}. Regarding both curves, the optimised value of α_{SVR2} will be equal to 2 as a higher value will significantly decrease the recall and a lower value will provide low efficiency.

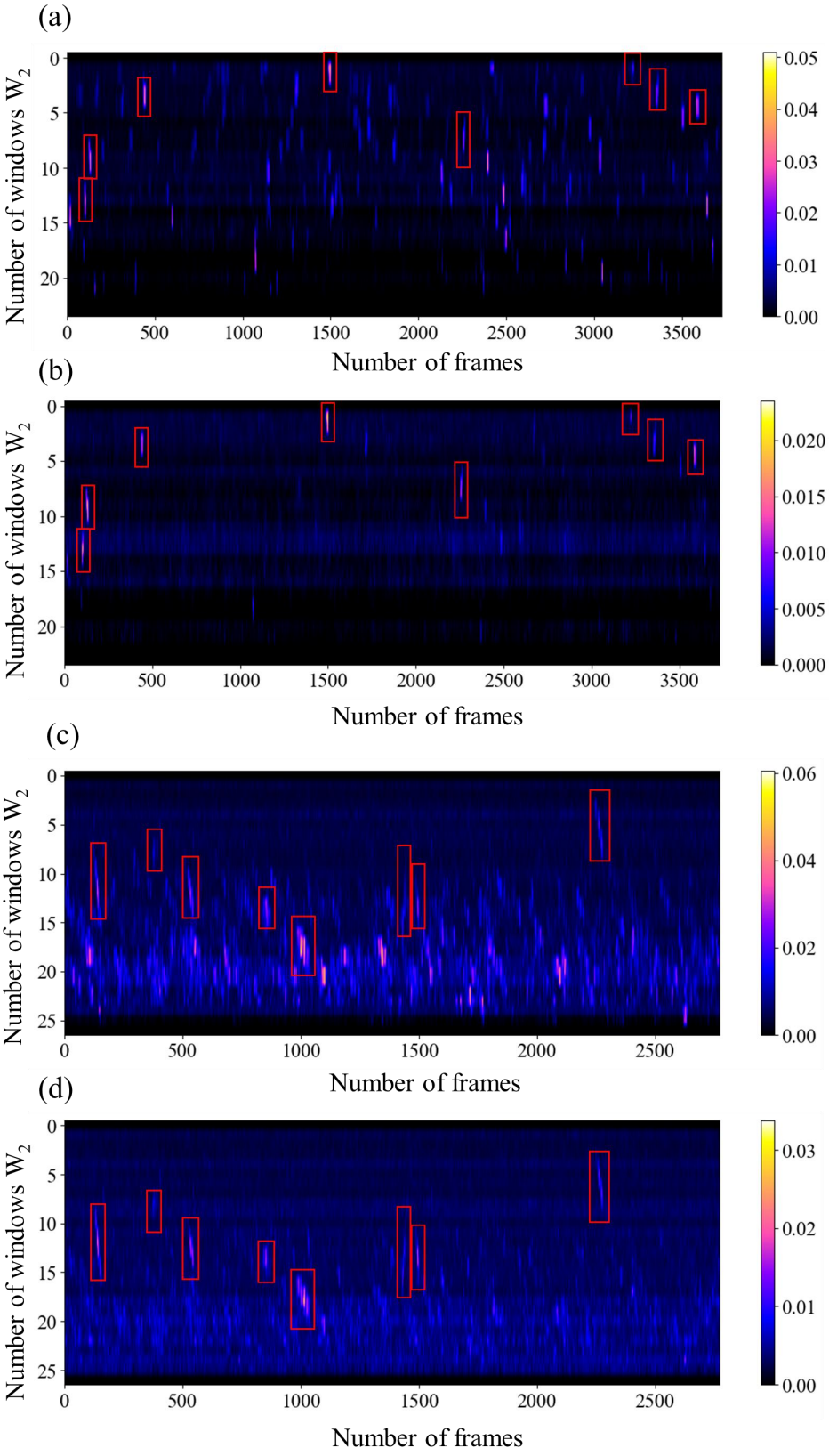


Figure 33 – Echograms calculated using the anguilliform-specific proposed method based on SVR, from a video of the MZC-ARIS dataset (with the 2^{nd} SVR (a) and the 3^{rd} SVR (b)) and a video from the MZC-BV dataset (with the 2^{nd} SVR (c) and the 3^{rd} SVR(d)).

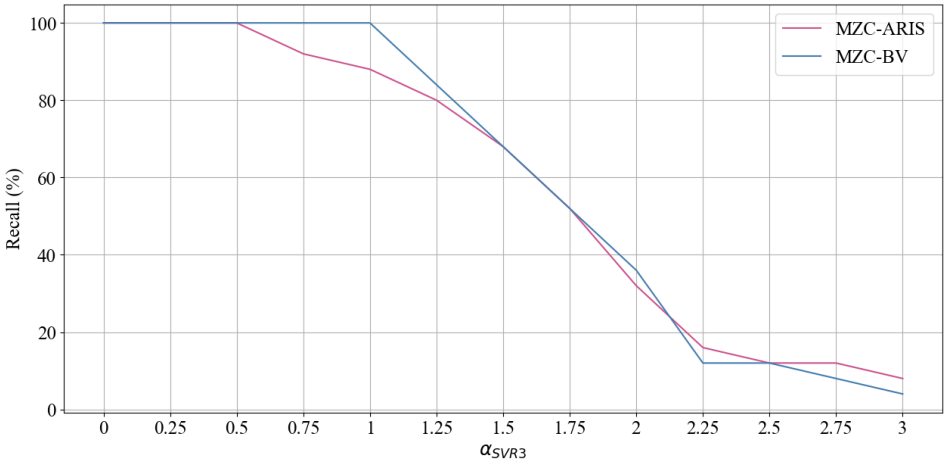


Figure 34 – Graph of the recall obtained with different values of α_{SVR3} for the MZC-ARIS dataset (in pink) and the MZC-BV dataset (in blue).

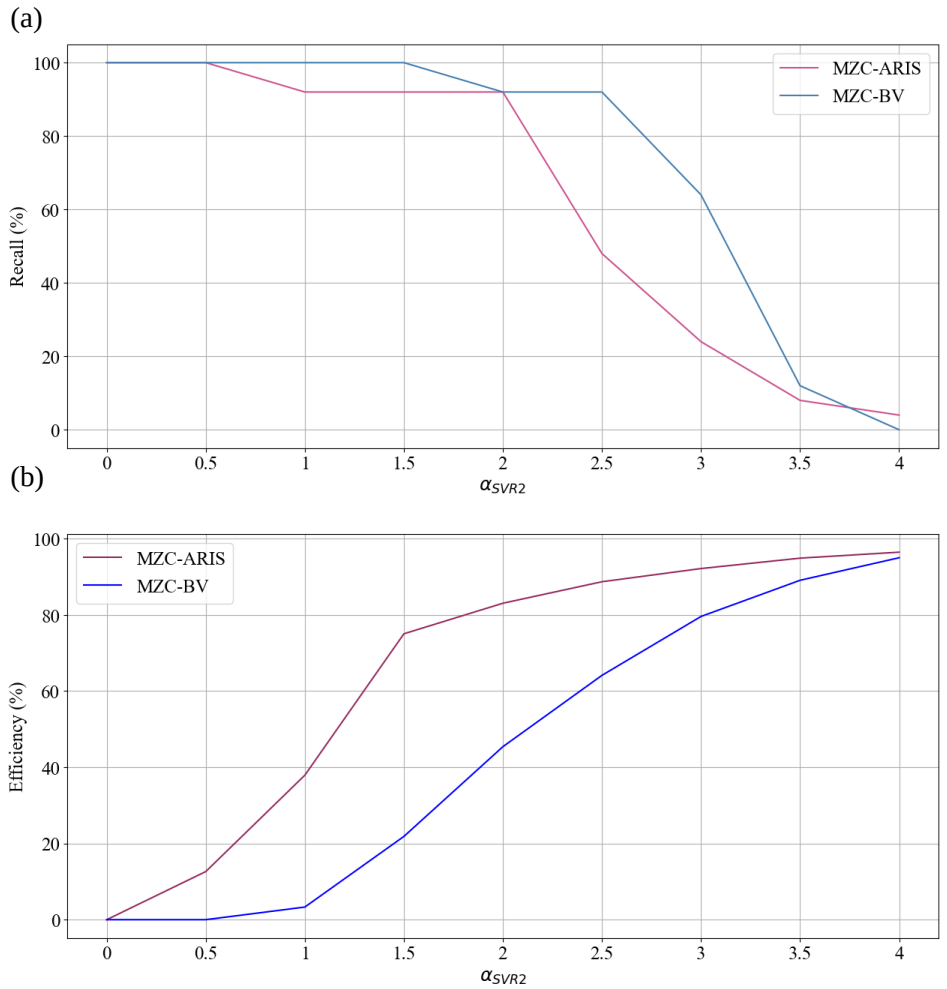


Figure 35 – Graph of the recall (a) and efficiency (b) obtained with different values of α_{SVR2} for the MZC-ARIS dataset (in pink) and the MZC-BV dataset (in blue).

Regarding the MZC-BV dataset, Figure 35.a displays a plateau of the recall until α_{SVR2} is equal to 2.5. A significant drop is then observed for larger values of α_{SVR2}. At the same time, Figure 35.b highlights a constant increase of the efficiency for values of α_{SVR2} superior or equal to 1. Regarding both curves, we choose α_{SVR2} equal to 2.5, as it allows us to get the highest efficiency possible without a drop of the recall.

The recall and efficiency reached with the tuned values of α_{SVR2} and α_{SVR3} are summarised in Table 4 for both MZC-ARIS and MZC-BV datasets.

Table 4 – Recall and efficiency obtained with the tuned anguilliform-specific method of FOI detection on MZC-ARIS and MZC-BV. Values of α_{SVR2} and α_{SVR3} are equal to 2 and 1 for MZC-ARIS respectively, while α_{SVR2} and α_{SVR3} are equal to 2.5 and 1.25 for MZC-BV respectively.

	Recall	Efficiency
MZC-ARIS (N = 25)	92%	83.1%
MZC-BV (N = 25)	92%	64.1%

Table 5 summarises the computation time required to calculate those echograms. It corresponds to the computation time for two videos from MZC-ARIS and MZC-BV respectively reached by paralleling the method on three logical processors (Computer characteristics: HP ZBook 15 G3, IntelCore i7-6700HQ CPU @ 2.60GHz, 2592MHz, 4 cores, 8 logical processors, 32 Go random-access memory). The computation time is equal to the MZC-ARIS videos duration while it is equal to more than 2.5 times the MZC-BV videos duration. The high computation time of this method is due to the repetitive call to the time consuming SVD function from *Numpy*. The BlueView frames having high width and height, the number of sliding windows required to go through each frame is higher than on ARIS videos. The call for the SVD function is therefore higher, increasing the computation time for the BlueView videos. This high computation time is an issue to the development of a real-time method of AC data analysis. Hence, the extraction of FOI is only the first step of the method but already requires a calculation time equal or much higher to the video duration.

Table 5 – Computation time of the anguilliform-specific method reached with paralleling on three logical processors (Computer characteristics: HP ZBook 15 G3, IntelCore i7-6700HQ CPU @ 2.60GHz, 2592MHz, 4 cores, 8 logical processors, 32 Go random-access memory). The computation time corresponds to those of two videos respectively from MZC-ARIS and MZC-BV respectively. The videos characteristics are displayed: the number of frames, the video duration and the frame height and width.

	Computation time	Number of frames	Video duration	Frame height (px)	Frame width (px)
MZC-ARIS	10' 3"	4194	10'	1276	664
MZC-BV	26' 47"	2775	10'	1238	2302

3.3 A generic approach of AC echogram calculation

The development of a second method aims to overcome the limitations of the anguilliform specific method. First, by reducing the computation time that is required for the AC echogram calculation to pursue the objective of a real-time AC data analysis. Second, by providing a generic approach that is not limited to anguilliform species.

3.3.1 Proposed method

A smoothing of the frame is firstly carried out using a median filter (eq. 4) [Ahmad and Sundararajan, 1987], with an empirically chosen large filter size (k = 100 mm) (Figure 36.b). The larger the filter size, the smoother the image is. The smoothed frames are then submitted to a Background Subtraction (BS) filter (Figure 36.c) to keep only the moving pixels. The BS filter used in the proposed approach is the Gaussian Mixture-based Background/Foreground Segmentation Algorithm based on [Zivkovic, 2004] and [Zivkovic and Van der Heijden, 2006]. It has been previously applied on acoustic videos [Fernandez Garcia et al., 2023] and can be easily implemented through the *createBackgroundSubtractorMOG2* function of the *OpenCV* library. The empirically chosen parameters values are the default one of 16 for the parameter *varthreshold* and a value of 15· Frames Per Second (FPS) for the *history* parameter, where FPS corresponds to the frame rate of the video.

Scanning of the binary image resulting from the BS algorithm can then be performed. Sliding windows scan each frame with 50% overlap along the vertical axis (Figure 37). The windows width is thus equal to the frame width while the window height depends, as for the anguilliform method, on the frame resolution r and on the target length L_{Target} (eq. 6).

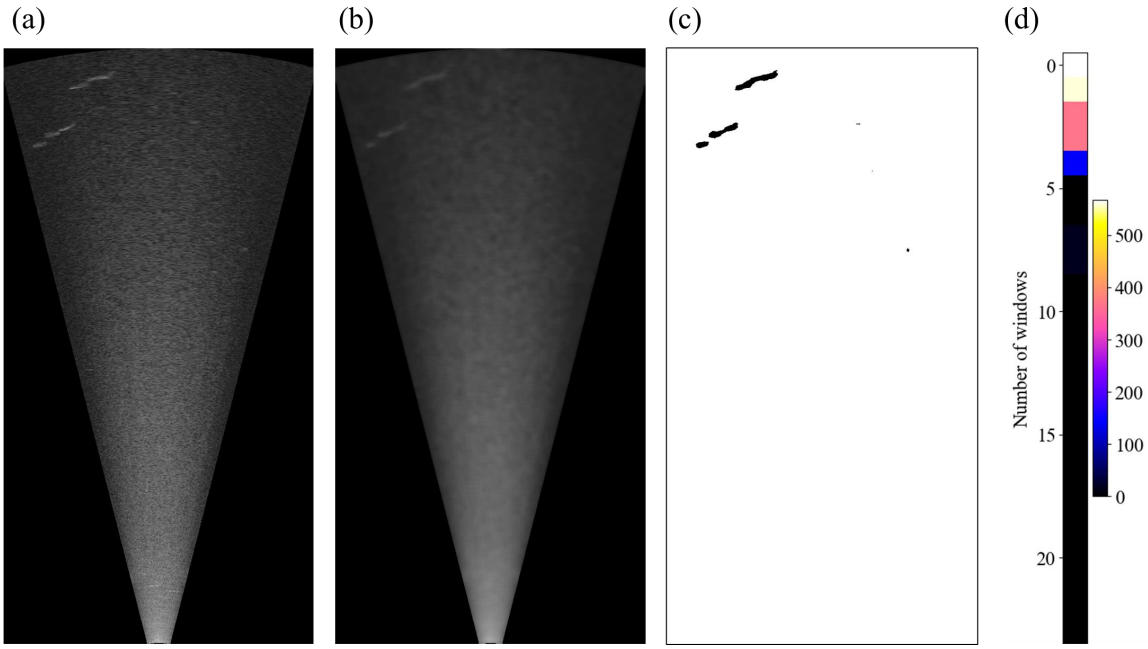


Figure 36 – Example of the pipeline deployment for the analysis of a frame in the echogram calculation process with the generic method. The input frame (a) is first smoothed using a median filter with a large structuring element (b), before a background subtraction is applied (c). The resulting scalar V summarising the largest cluster areas, expressed in square centimetres, per sliding window of the frame is calculated (d).

For each window, a clustering of the pixels is carried out and the area of each of them is retrieved using the *regionprops* function of the *skimage.morphology* package [Van der Walt et al., 2014]. The choice of large kernel size for the median filter enables us to merge nearby clusters prior to BS and thus to get the largest areas present in the frame. The maximum of the surface areas clustered in the windows (eq. 13) is kept as part of the resulting scalar V for the studied frame (Figure 36.d).

$$\forall w \in [1, W_2], V_{w,1} = max(A_{(I_{i,j})\ (w-1)*\frac{d_w}{2} \leq i \leq (w-1)*\frac{d_w}{2}+d_w \atop 1 \leq j \leq D_1}) \qquad \text{(eq. 13)}$$

With V, the scalar summarising the largest cluster present in each window; $A_{(I_{i,j})}$, the areas of each cluster present in the windows $I_{i,j}$.

Similarly to the anguilliform-specific method, these scalars progressively build up the echogram (Figure 38). The reading of the echogram is then carried out by thresholding the cluster's areas, using a threshold value α_{Area}. The value of α_{Area} chosen will therefore have an influence on the FOI extracted.

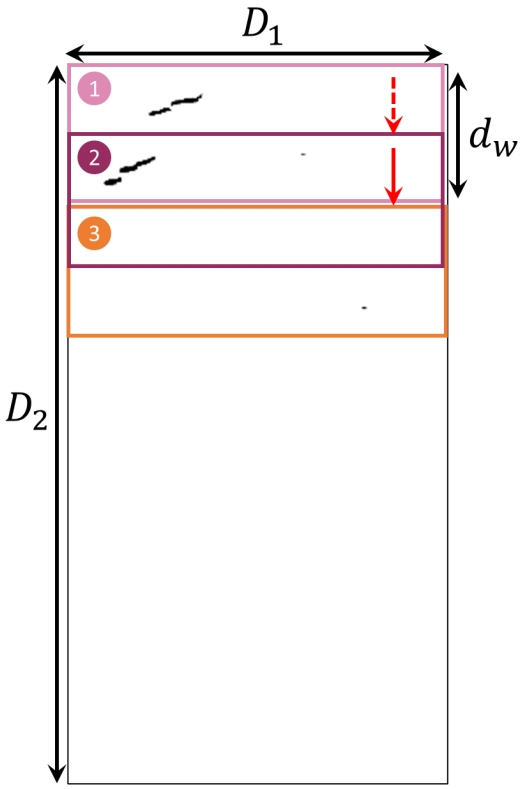


Figure 37 – Example of an acoustic frame after background subtraction, with windows sliding along the vertical axis, highlighted in colour. The two red arrows show the direction of the window for the two first slides. The width of the windows d_w as well as the size of the frame D_1 and D_2 are also indicated on the frame.

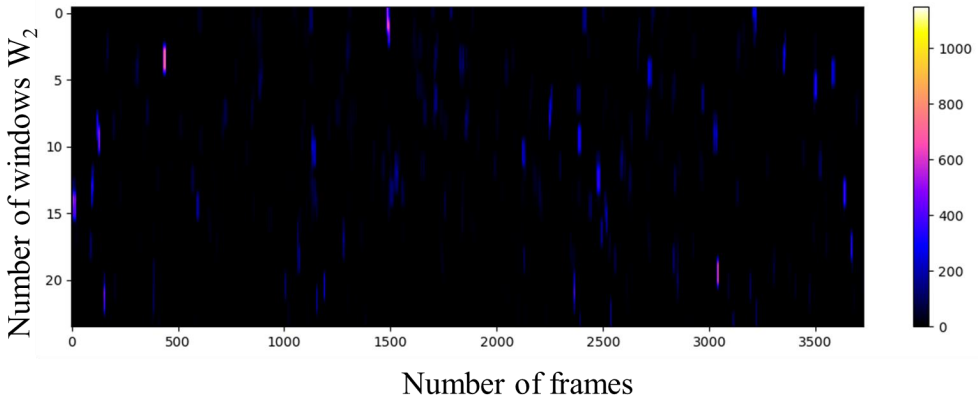


Figure 38 – Example of an echogram obtained using the generic method, on a 10-minutes video from the MZC-ARIS.

3.3.2 Metrics of performance

The tuning and evaluation of the method performance are carried out on 25 fish passages from the MZC-ARIS dataset, on 25 fish passages of the MZC-BV dataset and on 20 passages on the SEL-ARIS dataset. The 50 passages from MZC-ARIS and MZC-BV are the same that have been used to tune and evaluate the anguilliform-specific approach. The proposed generic

method aims to perform on all species, not only the anguilliform ones. The SEL-ARIS has thus been integrated into the analysis. For each of the three datasets, an example of the calculated echogram for one video will be displayed. Visual comparison with the value of N_{Fish} will thus be possible before tuning the threshold α_{Area} to enable automatic extraction of the FOI. Recall and efficiency will be evaluated according to the value of α_{Area}.

Computation time of the proposed approach is also investigated as another measure of performance.

3.3.3 Performance and tuning of our generic method

Figure 39 displays examples of echograms obtained using the generic method on the three datasets (MZC-ARIS, MZC-BV and SEL-ARIS). Red squares identify the passages of interest based on the operator truth. All of the fish passages counted by the operator can be seen on the three echograms. However, the values of displayed surface area vary from one passage to the other, some of them having lower values.

The automatic reading of these echograms is carried out using the threshold α_{Area} based on the objects surface areas expressed in square centimetres. Figure 40.a displays the recall values according to the values of α_{Area} while Figure 40.b displays the efficiency of the method depending on α_{Area}. According to the graphs, the optimal value for α_{Area} depends on the type of AC. On one hand, the recall curves for the MZC-ARIS and SEL-ARIS display a plateau of their values until α_{Area} equals to 200 cm^2 before a drop of the curves, with similar slope for both datasets. Regarding their efficiency values, a constant similar increase is observed before reaching a plateau from α_{Area} equals to 450 cm^2. From both curves, we can deduce an optimal value of α_{Area} equals to 200 cm^2. On the other hand, the recall curve of the MZC-BV dataset stays at 100% until it quickly decreases from α_{Area} equal to 300 cm^2. The efficiency curve highlights a slower increase compared to those of the two other datasets. A plateau is reached for α_{Area} around 600cm^2. All together, all of these observations lead to an optimal value of α_{Area} at 300 cm^2 for the videos recorded by the BV. The recall and efficiency reached with the tuned value of α_{Area} are summarised in Table 6 for each of the three datasets.

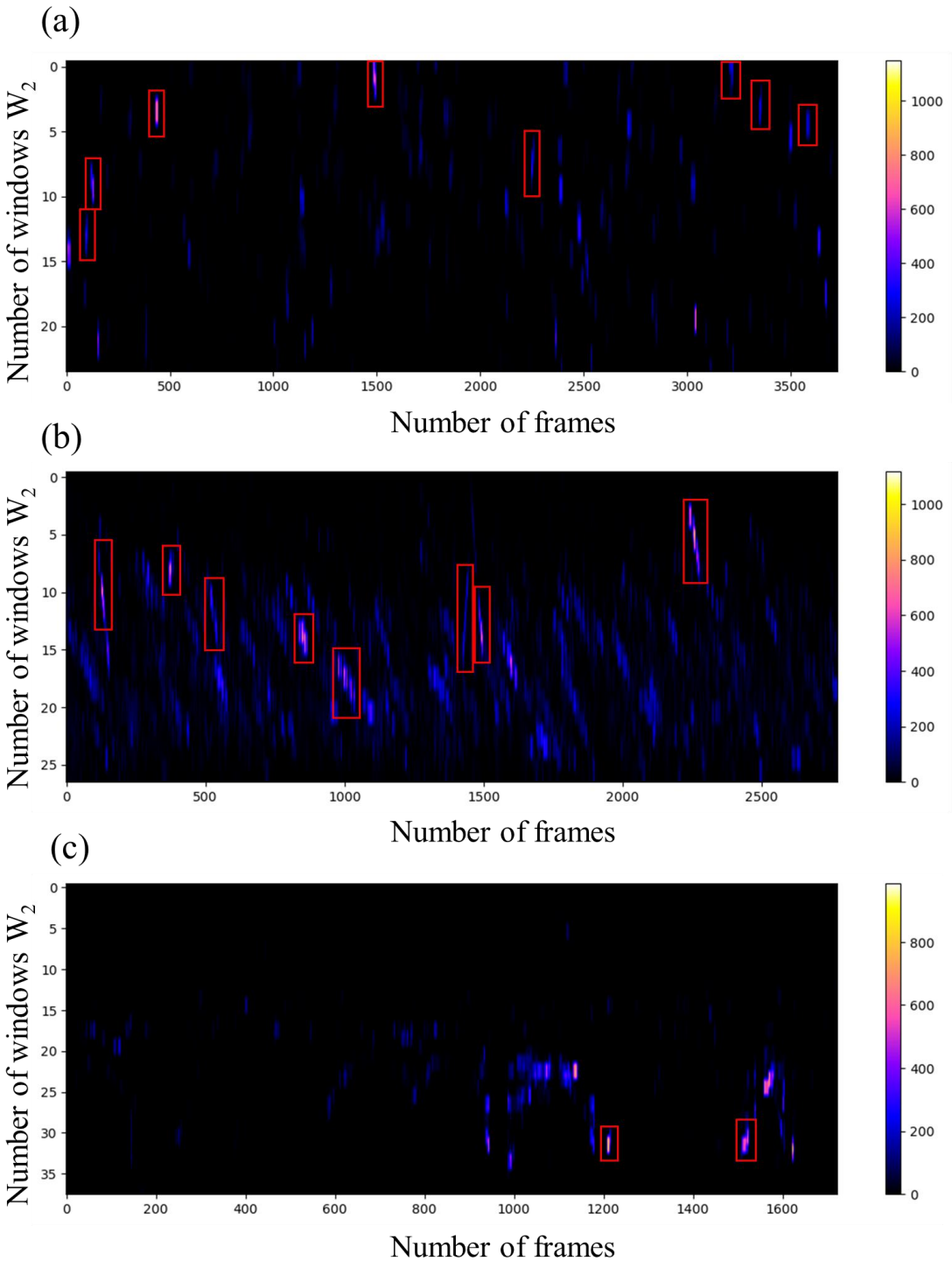


Figure 39 – Examples of echograms calculated using the generic proposed method based on objects surface areas, from a video of the MZC-ARIS dataset (a), a video from the MZC-BV dataset (b) and a video from the SEL-ARIS dataset (c). The colour bar represents the value of areas calculated, in cm^2.

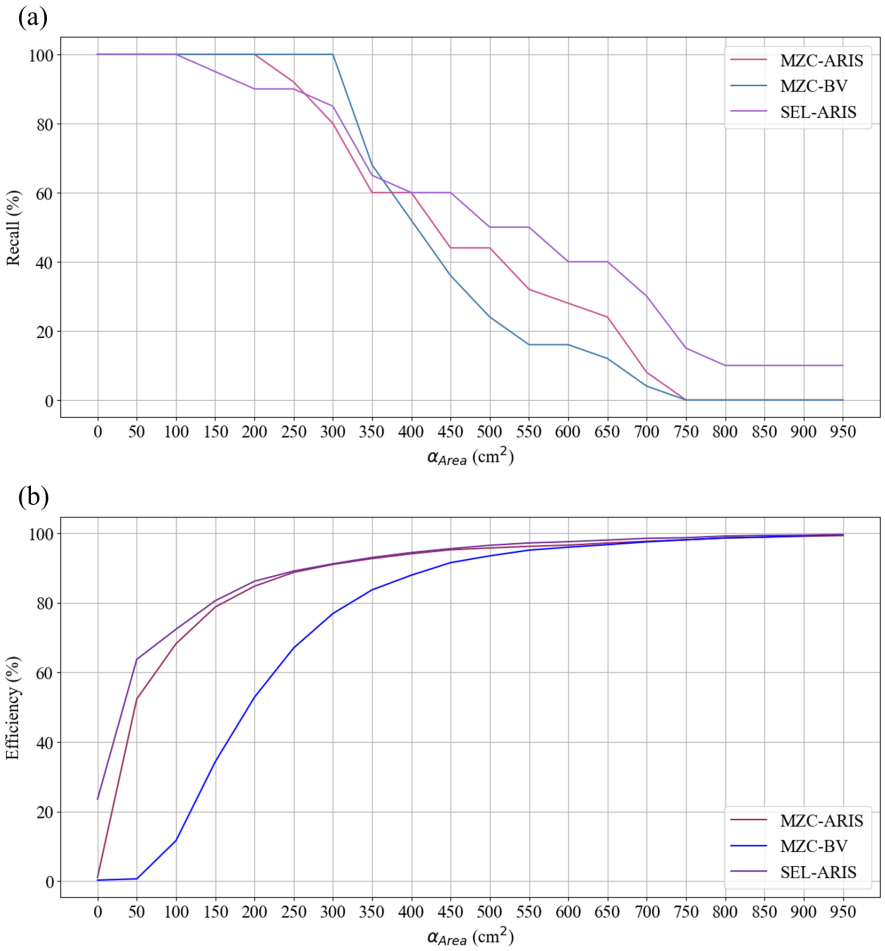


Figure 40 – Graph of the recall obtained with different values of α_{Area} for the MZC-ARIS dataset (in pink), the MZC-BV dataset (in blue) and the SEL-ARIS dataset (in purple) (a). Graph of the efficiency obtained with different values of α_{Area} for the MZC-ARIS dataset (in pink), the MZC-BV dataset (in blue) and the SEL-ARIS dataset (in purple) (b).

Table 6 – Recall and efficiency rates reached using the generic methods on the three datasets (MZC-ARIS, MZC-BV and SEL-ARIS). The value of α_{Area} used for the MZC-ARIS and SEL-ARIS datasets is equal to 200 cm^2 and the one used for the MZC-BV dataset is equal to 300 cm^2.

	Recall	Efficiency
MZC-ARIS (N = 25)	100%	84.7%
MZC-BV (N = 25)	100%	76.9%
SEL-ARIS (N =20)	90%	86.1%

Table 7 summarises the computation time required by the method. It corresponds to the computation time for three videos from MZC-ARIS, MZC-BV and SEL-ARIS datasets respectively, reached by paralleling the method on three logical processors (Computer characteristics: HP ZBook 15 G3, IntelCore i7-6700HQ CPU @ 2.60GHz, 2592MHz, 4 cores, 8 logical processors, 32 Go random-access memory). The proposed method requires a computation time lower than the video's duration.

Table 7 – Computation time of the generic method reached with paralleling on three logical processors (Computer characteristics: HP ZBook 15 G3, IntelCore i7-6700HQ CPU @ 2.60GHz, 2592MHz, 4 cores, 8 logical processors, 32 Go random-access memory). The computation time correspond to the ones of three videos respectively from MZC-ARIS, MZC-BV and SEL-ARIS datasets. The videos characteristics are displayed: the number of frames, the video duration and the frame height and width.

	Computation time	Number of frames	Video duration	Frame height (px)	Frame width (px)
MZC-ARIS	4' 36"	4194	10'	1276	664
MZC-BV	6' 26"	2775	10'	1238	2302
SEL-ARIS	2' 15"	1723	10'	1350	738

The comparison of the performance of the proposed generic method with the one shown in Table 4, highlights that the generic method shows similar efficiency and a slightly higher recall on MZC-ARIS dataset than the anguilliform-specific method. The efficiency of the anguilliform-specific method on MZC-BV dataset is 10 percentage points lower than the one reached by the generic method while the recall presents the same difference as for MZC-ARIS dataset. Besides, the anguilliform-specific method largely exceeds the computation time of the generic one.

Regarding the performance of both methods, the proposed generic method overcomes the limitations of the first developed anguilliform-specific method and will therefore be implemented in the whole proposed pipeline.

4 From frames of interest to regions of interest

The FOI being extracted from the acoustic videos, the next step of the process aims to automatically detect the image of the discontinuities, i.e., the ROI, using successive image processing techniques.

4.1 Image processing for object detection

Object detection is designed to distinguish the foreground pixels from the ones of the background, aiming to cluster pixels in ROI. Many methods exist in the literature [Piccardi, 2004, Cheung and Kamath, 2004, Herrero and Bescós, 2009] and are mainly developed to be adapted to specific cases [Kulchandani and Dangarwala, 2015, Tiwari and Singhai, 2017]. Methods differ depending on the nature of the video environment [Herrero and Bescós, 2009]. On the one hand, controlled environment data requires simple non-recursive methods such as frame differencing [Alex and Wahi, 2014] or median filter [Cheung and Kamath, 2004]. On the other hand, an uncontrolled environment may cause some noises, illumination change (change in pixel intensity) or viewpoints change [Herrero and Bescós, 2009]. Those case studies require more complex and recursive methods to efficiently detect the objects. The main recurrent methods [Piccardi, 2004, Herrero and Bescós, 2009] are the Mixture of Gaussians (MoG) [Stauffer and Grimson, 1999] and the Kernel Density Estimation (KDE) [Elgammal et al., 2002]. MoG uses a set of weighted Gaussian distributions to represent each pixel of the background and to estimate if a pixel belongs to the background according to its distance from the ordered distributions [Herrero and Bescós, 2009]. KDE is a non-parametric model that estimates at each frame the Probability Density Function (PDF) which is then used to determine if a pixel belongs to the modelled distribution and thus if it belongs to the background or not [Herrero and Bescós, 2009]. KDE and MoG are known to give a very good accuracy [Piccardi, 2004, Baf et al., 2007]. However, the main limitation of KDE is the time computation and memory issue considering that it needs to systematically estimate the PDF from the last values of each pixel of the frame [Herrero and Bescós, 2009]. MoG tends to overcome this memory issue using a background model that is progressively updated requiring less storage [Balaji and Karthikeyan, 2017, Baf et al., 2007]. Some drawbacks remain nonetheless for MoG, as it is still a computationally intensive method [Cheung and Kamath, 2004] although less than KDE [Baf et al., 2007] and it is sensitive to sudden changes in frame global illumination [Cheung and Kamath, 2004]. To overcome some of the issues encountered with MoG, [Zivkovic, 2004] and [Zivkovic and Van der Heijden, 2006] proposed an updated method, MoG2. Contrary to MoG, MoG2 enables the selection of the appropriate number of Gaussian distributions to represent each pixel [Zivkovic and Van der Heijden, 2006]. It aims to overcome the illumination changes issue of MoG and provide a better accuracy of the background subtraction [Lee and Park, 2012, Cheong et al., 2019], while reducing the computation time [Zivkovic and Van der Heijden, 2006].

4.2 Detection of ROI

A first processing, similar to the one of the generic echogram calculation, is applied on the frame (Figure 41.a): a median filter, with smaller structuring element (empirically set up at 50 mm), is applied to get a thin smoothing of the frame (Figure 41.b) before applying the BS filter and obtaining the image *BS* (Figure 41.c).

The background subtraction step often leads to fragmented or broken binary images (Figure 41.c). To overcome this issue and try to avoid detecting multiple times a same ROI, we dilate the image (Figure 41.d) by applying the corresponding mathematical morphology operation (eq. 14). A structuring element is chosen deliberately large to overcome this issue, corresponding to a rectangle of dimension 15x15 pixels.

$$\forall\, u \in [1, D_1],\ \forall\, v \in [1, D_2],\ I_{\oplus}(u, v) = max(I(u - x : u + x, v - y : v + y)) \qquad \text{(eq. 14)}$$

With I the input frame; $I_{\oplus}$, the dilated frame; D_1 and D_2, the size of the frame, in pixels; and, (x,y), the size of the structuring element.

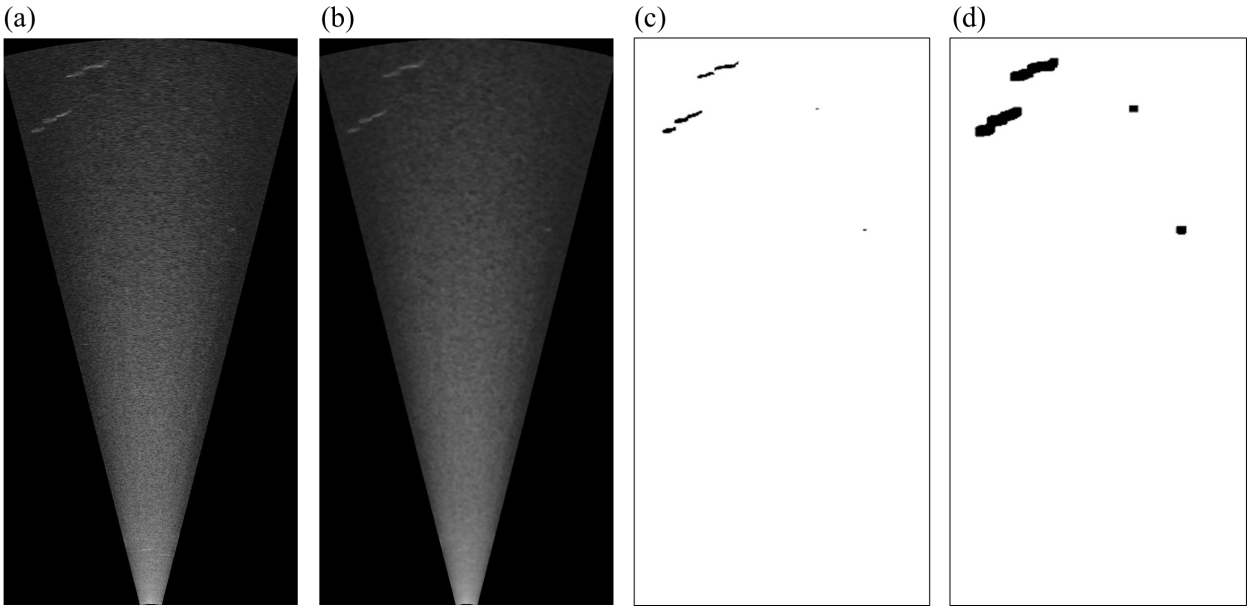


Figure 41 – Pipeline for the analysis of a frame in the ROI detection step. The input frame (a) is first smoothed using a median filter (b), before a background subtraction is being applied (c). A dilation of the binary image is then performed in order to merge close regions and detect the ones of interest (d).

A filter of the resulting ROI R is then carried out to remove the smallest areas that cannot correspond to targeted fish (eq. 15). The filtering is done on the main axis length of the ROI, i.e., the length of the main axis of the ellipse that fits the ROI. This property gives an indication of the length of the ROI with the advantage of being able to be quickly calculated using the

regionprops function of *scikit-image* package [Van der Walt et al., 2014]. Regions whose major axis length is greater than the threshold ϵ_L are kept as ROI and extracted in a square vignette (Figure 42). We choose ϵ_L equal to 25% of L_{Target}.

$$\forall\, a \in [1, N_R],\ R_F(a) = \begin{cases} R(a) & \text{if } L(a) \geq \epsilon_L \\ 0 & \text{else} \end{cases} \qquad \text{(eq. 15)}$$

With N_R, the number of ROI before filtering; R the ROI before filtering; R_F, the result ROI after filtering; L, the main axis length of each ROI of R; and ϵ_L, the threshold used to carry out the filtering.

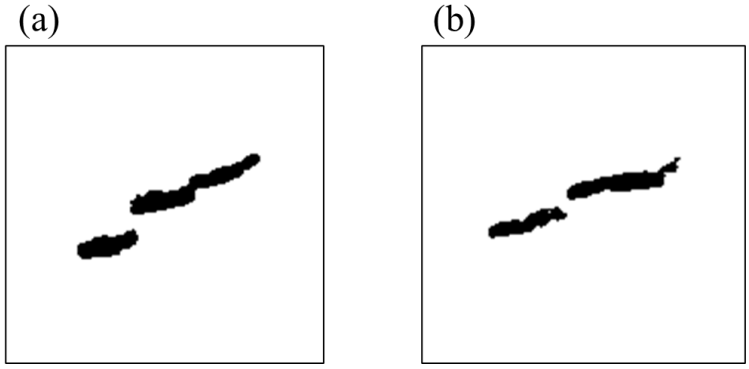


Figure 42 – The two vignettes of the ROI extracted after the process displayed in Figure 41. Each vignette represents one single target.

4.3 Restoration of ROI

As already mentioned, and illustrated in Figure 42, ROI images are often fragmented. To maximise the image information for the following steps of the process, we propose a restoration of the ROI images (Figure 43.a) through a succession of mathematical morphological filters. The vignette *BS* is extracted after the BS filtering (Figure 43.b). From there, we remove small noise through a closing of the image (Figure 43.c) using a small structuring element K_F of 5x5 mm, leading to the F image (eq. 16).

$$F = BS \bullet K_F \qquad \text{(eq. 16)}$$

An oriented dilation (eq. 17) is then applied to obtain the image D (Figure 43.d). The structuring element K_O depends on the global orientation of the ROI. The structuring element shape is set to follow as much as possible the orientation of the ROI. If the orientation is closer to the horizontal axis, then K_O corresponds to a lying rectangle of 50x120 mm; if the orientation is closer to the vertical axis, then K_O corresponds to a standing rectangle of 120x50 mm; and, if the orientation is closer to the diagonal axis, then K_O corresponds to a square of 120x120 mm. Figure 44 displays an example of a ROI of 65° orientation, dilated with the three

possible structuring elements, highlighting the benefit of an oriented dilation to avoid taking into account close noise.

$$D = F \oplus K_O \quad \text{(eq. 17)}$$

The free-of-noise resulting image I (Figure 43.e) corresponds to the intersection of the BS image with the D one (eq. 18).

$$I = BS \cap D \quad \text{(eq. 18)}$$

Finally, the restoration is carried out through a dilation with a structuring element K_D of 120x120mm corresponding to MxM pixels according to the frame resolution r (M $= \frac{120}{r}$), followed by an erosion K_E of M-1xM-1 pixels (eq. 19) allowing us to retrieve the final image R (Figure 43.f).

$$R = ((I \oplus K_D) \ominus K_E) \quad \text{(eq. 19)}$$

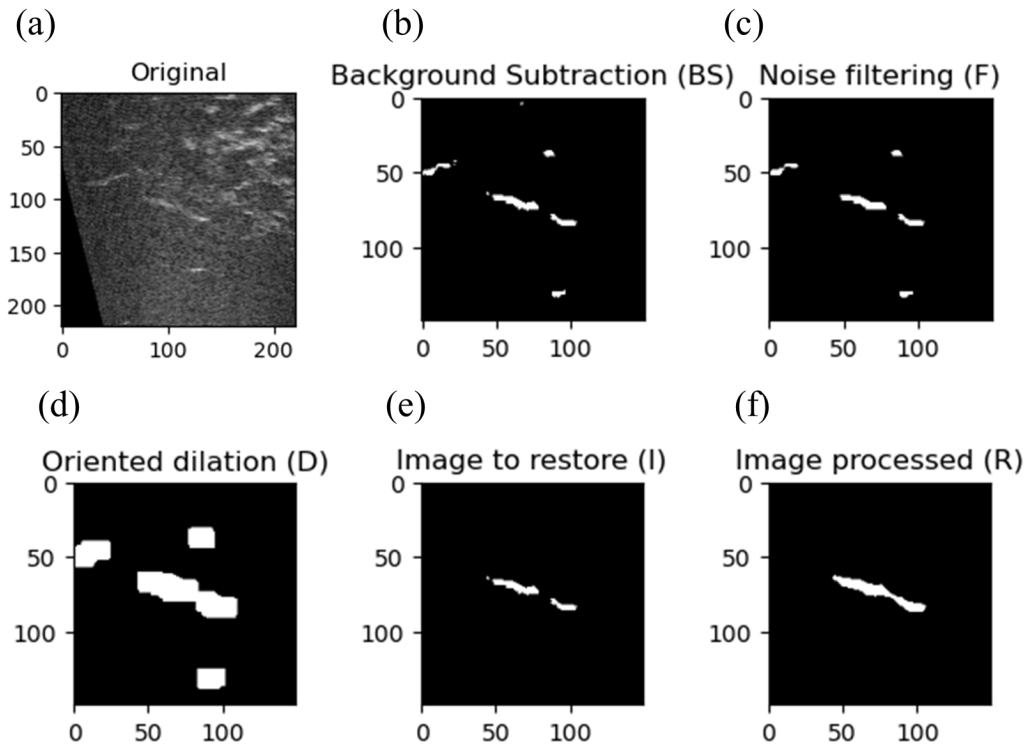


Figure 43 – Pipeline of the restoration of a ROI image. The original image (a) is first processed using the pipeline displayed in Figure 41. The restoration starts from the image after BS (b). A noise filtering is carried out through a closing (c) before applying an oriented dilation (d) to the resulting binary image. The image to restore (e) is calculated as the intersection of the BS image (b) with the oriented dilation (d). The restoration can finally be applied on this image through a dilation of the image followed by an erosion (f).

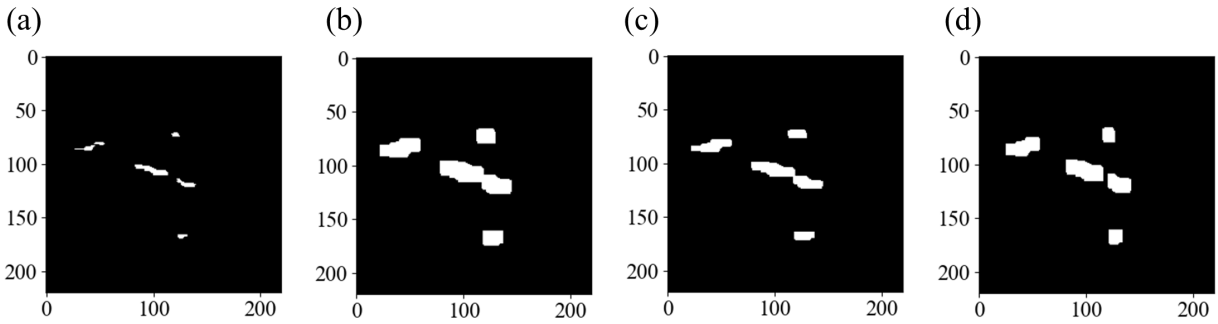


Figure 44 – Example of a noise-free vignette (a), dilated with three different structuring elements, a square (b), a lying rectangle (c) and a standing rectangle (d). The orientation of the ROI is equal to 65° and it is thus closer to the diagonal axis (45°). Our oriented dilation method will thus be carried out using a squared structuring element. The three dilated images displayed highlight the coherence of the choice of structuring element as well as its importance.

5 From regions of interest to targets

From all of the ROI present in a frame at t=k, we want to detect the targets, meaning we want to track our ROI along their respective movement in the camera FOV. The successive detections of a same ROI define one target.

5.1 Image processing for object tracking

Tracking methods aim to follow-up objects along their entire movements in the camera FOV, by carrying out frame to frame analysis. Methods differ depending on the characteristics they use to perform the tracking process: on one side the kernel-based methods, on the other side the silhouette-based ones and finally the point-based methods. Several reviews [Athanesious and Suresh, 2010, Balaji and Karthikeyan, 2017, Tiwari and Singhai, 2017] discussed the advantages and drawbacks of these methods and thus help us to discuss the best one to use depending on the case study. Kernel-based methods first describe the input object through geometric shapes before performing the tracking process on those shapes [Comaniciu et al., 2003]. Some limits have been identified with these methods as shapes might not be fully able to encapsulate the object and background elements might be considered as pieces of the object [Balaji and Karthikeyan, 2017]. Silhouette-based methods [Rosenhahn et al., 2005] might overcome this issue as they used a model to characterise either contour or shape of the object [Balaji and Karthikeyan, 2017], allowing a more accurate, but also a more time consuming, description and tracking process. Using higher scale characteristics, point-based methods track an object based on a characterising point, usually its centroid. Very basic algorithm consists of calculating, from one frame to another, the Euclidean distance between two objects' centroids and choosing the closest point as the continuity of the tracking. Despite the very easy implementation, computation time and memory storage it requires, performance of this method might decrease in case of multi-

targets tracking. More robust methods are available to handle multi-target tracking despite more complex implementations. We found three main algorithms in the literature [Balaji and Karthikeyan, 2017, Tiwari and Singhai, 2017] that carried out those point-based tracking. The particle filter [Hue et al., 2002] corresponds to an optimisation function where the solution is searched using particles. A particle corresponds to a position. For each of them, a test is run to calculate the probability for the tracked point to be located at this position, solving the optimisation function. It enables us to model non-linear object motion but presents the disadvantage of being costly in computation time; a high quantity of particles being required to reach good performance [Athanesious and Suresh, 2010]. Another common point-based algorithm is the Multiple Hypothesis Tracking (MHT) [Cox and Hingorani, 1996]. However, it presents the same issue in terms of computation time as the particle filter, in addition to high needs of memory [Blackman, 2004]. The last method is the Kalman filter. The algorithm estimates unknown variables from a series of measurements made over time [Welch and Bishop, 2006]. The method is applicable even in noisy images [Tiwari and Singhai, 2017], and usually requires lower computation time than particle filter or MHT algorithm [Tiwari and Singhai, 2017].

5.2 Methods

5.2.1 Tracking using centroid-to-centroid distance

One of the simplest tracking methods is based on the centroid-to-centroid distance. Each track is composed of successive detections. For each of them, the centroid coordinates are calculated and saved in a scalar χ_n for the nth track studied (eq. 20). To link the track with its detections, Euclidean distance is calculated between track's centroid $\chi_{n,k-1}$ at t=k-1 and the centroid z_k of the observation at t=k, i.e., the centroid z_k of one of the ROI detected in the frame at t=k (Figure 45).

$$\forall k \in [1, K], \chi_{n,k} = (x_{n,k}, y_{n,k}) \quad \text{(eq. 20)}$$

With K, the number of detections in the track χ_n; and, $(x_{n,k}, y_{n,k})$, the centroid coordinates of the detection at the time step t=k.

The aim of the method is to match z_k with the closest track's centroid (eq. 21) as long as the Euclidean distance between both centroids does not exceed the distance error parameter ϵ_D (eq. 22). We defined ϵ_D dependent on L_{Target}. Hence, if a small fish is tracked, ϵ_D must be lower than in the case of large fish tracking. We thus choose ϵ_D equal to 1.5 * L_{Target}.

$$j = argmin_{i \in [1,N]}(\|z_k - \chi_{i,k-1}\|_2) \quad \text{(eq. 21)}$$

$$\|z_k - \chi_{j,k-1}\|_2 < \epsilon_D \quad \text{(eq. 22)}$$

With N, the number of tracks.

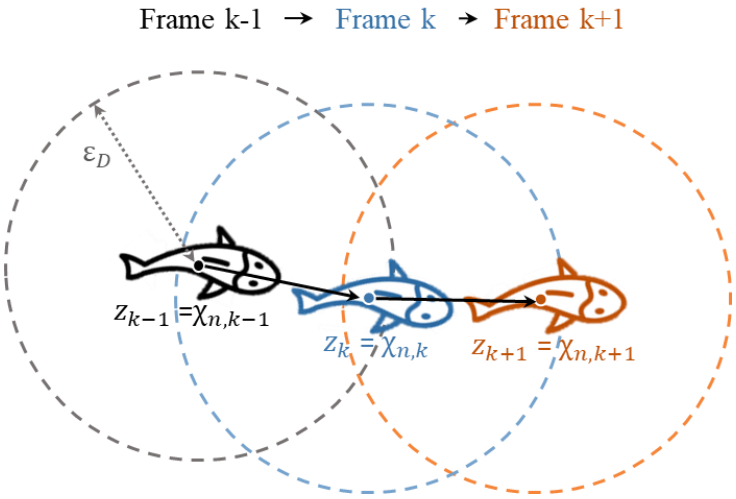


Figure 45 – Scheme of the principle of tracking using centroid-to-centroid distance for three successive detections represented in black, blue and orange. The circles correspond to the search area.

This constraint on distance is completed by a second one on the track's orientation Φ and observation's orientation θ_k. These orientations correspond to the angle of the target's body major axis with the horizontal axis (Figure 46). The aim of this second constraint is to keep the consistency of the orientation from one time step to the other. We assume here that the sampling period is short enough so that the orientation of the target cannot vary more than a maximum angle of ϵ_θ (eq. 25), arbitrarily set at 25°.

$$|\Phi_{j,k-1} - \theta_k| < \epsilon_\theta \qquad \text{(eq. 23)}$$

Figure 46 – Description of the orientation calculated for each observation and used for the tracking process.

In case of missing observation, meaning that at t=k, for one track χ_n no z_k matches the position of its last centroid $\chi_{n,k-1}$, then this track is updated as:

$$\chi_{n,k} = \chi_{n,k-1} \qquad \text{(eq. 24)}$$

$$\Phi_{n,k} = \Phi_{n,k-1} \qquad \text{(eq. 25)}$$

This update might raise issues as the distance between $\chi_{n,k+1}$ and z_{k+1} will be higher in the case of missing observations than if one observation would have been made. A last constraint has thus to be settled. A delimited time period ϵ_T is defined. It represents the maximum time period above which z_k cannot be considered as the continuity of one of the tracks χ_n even if it meets the first two conditions on distance and orientation. We choose empirically ϵ_T equal to 5 frames as a balance between allowing missing detections and avoiding to merge two individuals in a same track.

This first method might be challenged in the configuration of multi-targets or in case of missing observation and is therefore combined with a Kalman tracking filter.

5.2.2 Tracking using Kalman filter

The Kalman filter is an algorithm that estimates the values of measured variables over time [Kim and Bang, 2018]. It means that it allows us to predict the state of a process based on its previous values. The Kalman filter process can be summarised as (Figure 47):
1. Estimation of the current state $\tilde{\chi}_{k+1}$ from the previous state $\hat{\chi}_k$
2. Correction $\hat{\chi}_{k+1}$ of the estimation regarding the measured observation z_{k+1}
Steps 1 and 2 are then repeated at each iteration of the process.

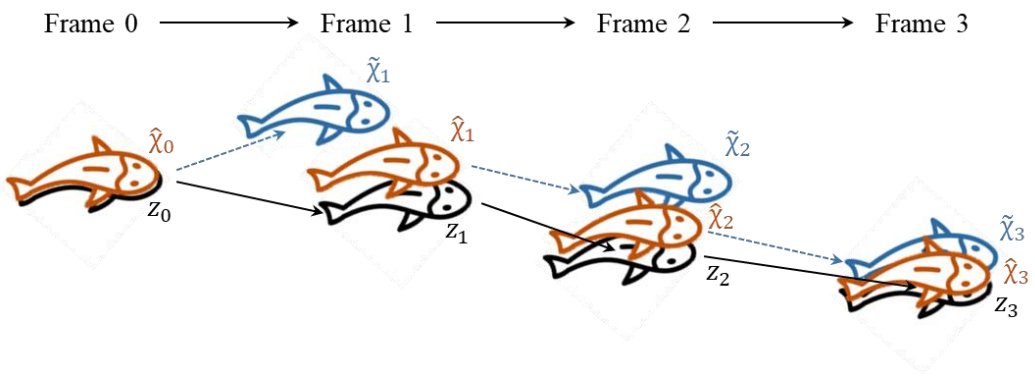


Figure 47 – Scheme of the Kalman tracking process for the first four predictions of a tracking. The dark fish correspond to the observed positions of the fish. The blue fish correspond to the successive estimated positions of the fish. The orange fish correspond to the successive corrected positions of the fish.

Regarding our tracking issue, we design our dynamic model as a position-velocity model, meaning that velocity is supposed not constant and estimated at each iteration as for positions. Our process state χ is thus defines as:

$$\chi = \begin{pmatrix} x \\ y \\ v_x \\ v_y \end{pmatrix} \quad \text{(eq. 26)}$$

With (x,y) and ($\dot{x}$,$\dot{y}$), the positions and the velocities along the vertical and horizontal axis respectively.

The state χ follows the usual kinematic equations:

$$x_k = x_{k-1} + \dot{x}_{k-1}T \quad \text{(eq. 27)}$$

$$\dot{x}_k = \dot{x}_{k-1} + \ddot{x}_{k-1}T \quad \text{(eq. 28)}$$

The predicted state $\tilde{\chi}_k$ and the corrected state $\hat{\chi}_k$ are calculated following (eq. 29) and (eq. 30) respectively, from [Welch and Bishop, 2006].

$$\tilde{\chi}_k = \Phi\hat{\chi}_{k-1} \quad \text{(eq. 29)}$$

$$\hat{\chi}_k = \tilde{\chi}_k + K_k(z_k - H\tilde{\chi}_k) \quad \text{(eq. 30)}$$

With z_k, the observation at t=k; H, the measurement matrix; K_k, the Kalman gain; and, Φ, the transition matrix from t=k-1 to t=k in our position-velocity system, defined as:

$$\Phi = \begin{pmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{(eq. 31)}$$

The Kalman filter being a recursive one, the state must be initialised at t=0. In our Kalman filter, we set up the initialisation based on the first two detections of the current track. It means that the track is started using the centroid-to-centroid distance method. After the first two detections, the corresponding Kalman filter is initialised and then used for the following detections. The state χ at t=0 is so defined as:

$$\tilde{\chi}_0 = \hat{\chi}_0 = \begin{pmatrix} z_{1,x} \\ z_{1,y} \\ \frac{z_{1,x} - z_{0,x}}{T} \\ \frac{z_{1,y} - z_{0,y}}{T} \end{pmatrix} \quad \text{(eq. 32)}$$

From (eq. 30), the Kalman gain K_k is defined by:

$$K_k = \tilde{P}_k H^T (H \tilde{P}_k H^T + R)^{-1} \quad \text{(eq. 33)}$$

The matrix P corresponds to the estimation error covariance matrix estimated following (eq. 34) and corrected following (eq. 35).

$$\tilde{P}_k = \Phi \hat{P}_{k-1} \Phi^T + Q \quad \text{(eq. 34)}$$

$$\hat{P}_k = \tilde{P}_k - K_k H \tilde{P}_k \quad \text{(eq. 35)}$$

The matrix R of (eq. 33) is the measurement noise covariance matrix and the matrix Q of (eq. 34) is the process noise covariance matrix. H, R and Q are adjusted from [Saho, 2018] to fit the 2D position-velocity system. They depend on the sampling period T, on the variance of the position measurement errors B_x, on the variance of the velocity measurement errors B_v and on the process noise σ_q.

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \quad \text{(eq. 36)}$$

$$R = \begin{pmatrix} B_x & 0 \\ 0 & B_v \end{pmatrix} \quad \text{(eq. 37)}$$

$$Q = \begin{pmatrix} \frac{T^4}{4} & 0 & \frac{T^3}{2} & 0 \\ 0 & \frac{T^4}{4} & 0 & \frac{T^3}{2} \\ \frac{T^3}{2} & 0 & T^2 & 0 \\ 0 & \frac{T^3}{2} & 0 & T^2 \end{pmatrix} * \sigma_q^2 \quad \text{(eq. 38)}$$

From the previous equations, three parameters B_x, B_v and σ_q must be set up to define the Kalman filter. We tune these parameters on a subset of our data to maximise the precision of the state prediction and estimation. The subset used was composed of 35 passages of fish of three different species, European eel, Atlantic salmon and European catfish. 180 combinations of the three parameters have been tested, with $B_x, B_v \in [0.01, 0.05, 0.1, 0.5, 1, 5]$ and $\sigma_q \in [0.5, 1, 5, 10, 15]$. For each configuration and each passage, the mean square error MSE (eq. 39) and the variance of the distance VAR (eq. 40) between prediction state and observation were calculated.

$$MSE = \frac{1}{n} \sum_{k=1}^{n} (z_k - \tilde{\chi}_k) \quad \text{(eq. 39)}$$

$$VAR = \frac{1}{n} \sum_{k=1}^{n} (d_k - \bar{d}) \quad \text{(eq. 40)}$$

With n, the number of detections per passage.

Our aim is to get the configuration of the Kalman filter that minimises both the MSE and the VAR, we therefore want to minimise the ratio:

$$r_K = \frac{1}{2}\left(\frac{MSE}{min(MSE)} + \frac{VAR}{min(VAR)}\right) \quad \text{(eq. 41)}$$

Figure 48 displays this ratio for each of the tested combinations with a maximum display threshold of 1.1, allowing us to clearly read the best combination that meets our objective. The closer to 1 r_K is, the darker the colour blue is displayed on Figure 48. The combination that reaches the ratio closer to 1 is therefore $B_x = 0.1$, $B_v = 0.5$ and $\sigma_q = 10$.

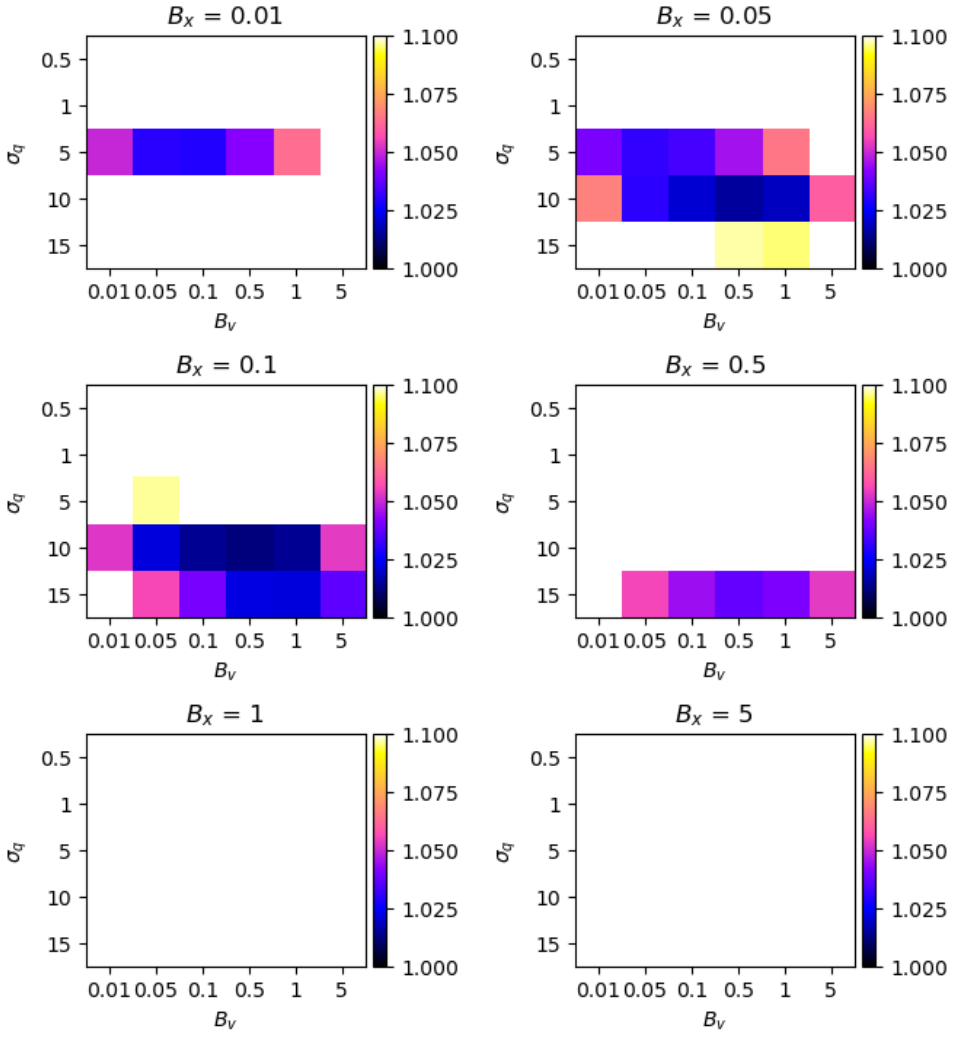


Figure 48 – Ratios r_K calculated on the development dataset for each combination of the Kalman filter parameters B_x, B_v and σ_q.

5.3 Performance of the tracking methods

5.3.1 Metrics of performance

Both tracking methods have been applied on videos from the MZC-ARIS and MZC-BV datasets. Those data are composed of European eel passages and are mostly isolated ones. Besides, both datasets have been recorded in a passage location with fish passing straight in the camera FOV. Regarding the centroid-to-centroid tracking method, the performance has been evaluated on 753 passages of fish from the MZC-ARIS dataset and 198 passages of MZC-BV dataset, as part of the study carried out by [Le Quinio et al., 2023]. Regarding the Kalman filter tracking method, the performance has been evaluated on 15 passages of fish from the MZC-ARIS dataset and 15 passages of fish from the MZC-BV dataset.

The automatic counting of both methods is compared to the operator reference and the corresponding precision value (eq. 42) will be calculated. In our case, precision value aims to evaluate if all fish passages have been counted only once. It is calculated from the True Positives (TP) and the False Positives (FP). All counting of the same fish passage more than once is considered as a tracking error and consequently as a FP. A tracking error corresponds to a same eel counted more than once because they are considered as multiple independent individuals by the tracking algorithm.

$$Precision = \frac{TP}{TP + FP} \qquad \text{(eq. 42)}$$

The performance evaluation of the centroid-to-centroid tracking method is part of the analysis of the pipeline of [Le Quinio et al., 2023], aiming to count and identify anguilliform fish. To evaluate its performance, sources of errors have been identified. To do so, an operator has checked each passage identified by the pipeline and categorised the errors according to their nature. The complete analysis is presented in Chapter 3. Here I will only focus on the performances of the tracking step to evaluate the centroid-to-centroid tracking approach.

5.3.2 Performance using centroid-to-centroid distance

Table 8 summarises the precision values reached by each dataset with and without the tracking errors, highlighting the impact of those errors and the performance improvement that can be made by fixing them. Tracking errors represent 6.4% of all of the errors identified on the MZC-ARIS dataset and 50% on the MZC-BV dataset. Consequently, we observe a small increase of the precision in the case the tracking errors are solved for the ARIS dataset (1 percentage point), while the increase is much more significant for MZC-BV with a precision increased by 11 percentage points.

Table 8 – Precision of the pipeline proposed by [Le Quinio et al., 2023] on two datasets (MZC-ARIS and MZC-BV). The percentage of errors due to incorrect tracking is highlighted as well as the precision that could have been obtained without these errors.

	MZC-ARIS	MZC-BV
Number of passages	753	198
Number of tracking errors	7	24
Percentage of tracking errors	6.4%	50%
Precision of the pipeline of [Le Quinio et al., 2023]	83.4%	74.3%
Precision of the pipeline of [Le Quinio et al., 2023] without tracking errors	**84.4%**	**85.3%**

5.3.3 Performance using the Kalman filter

To evaluate the effect of the Kalman filter implementation on the tracking process, it has been applied on a total of 30 passages of eels for which the performance of the centroid-to-centroid tracking method has also been assessed. The centroid-to-centroid tracking method has efficiently tracked 9 passages from the MZC-ARIS dataset and 7 passages from the MZC-BV dataset and therefore incorrectly tracked 6 and 8 passages respectively. By applying the Kalman filter on those passages, all have been correctly tracked and counted as one single passage, reaching a 100% precision rate for both MZC-ARIS and MZC-BV.

6 Discussion

The method described and evaluated in this chapter aims to reduce the amount of data to process in order to only focus on sequences of interest. From the FOI extraction to the target detection while including the ROI clustering, each step has an influence on the final identification result. The choice in their design, the optimization of some thresholds and their final performance have thus to be evaluated and optimised.

The proposed pipeline uses some of the steps already known in acoustic imaging analysis, in particular those essential for ROI detection and tracking. No matter the aims of the analysis, both steps are systematically processed in the automatic analysis of acoustic videos. [Wei et al., 2022] provides an interesting review of all of the image processing tools applied on acoustic videos to carry out those processing steps.

6.1 Detection of target passages on acoustic videos

The anguilliform-specific and the generic methods of echogram calculation proposed in this work allow us to get a clean and individual signal of target passages on the echograms enabling an automatic and simple reading by thresholding. The evaluation of both methods performance highlights almost similar performance on the ARIS data. However, on the BlueView videos, the generic method outperforms the anguilliform one in addition to reaching lower computation time suggesting the possibility of a real-time application. Regarding the anguilliform method, the use of SVD, a time-consuming mathematical operation, explains the higher computation time. Different SVD functions from the *numpy* and *pytorch* Python packages have been tried in vain. In light of these conclusions, the generic method of echogram calculation is included in our final pipeline. However, it must be noticed that the thresholding value allowing the reading of the echogram is device-dependent, with the value of α_{Area} equals to 200 cm^2 for the ARIS devices and the value of α_{Area} equals to 300 cm^2 for the BlueView ones. An identification of the AC types will therefore be made at the beginning of the pipeline based on the format of the acoustic videos. Besides, the choice of these thresholds was made from a limited number of passages. Although giving encouraging performances, it would be relevant to verify these choices on other sets of targets passages.

In addition to its performance, the use of the surface area values from the background subtraction for the generic method limits the impact of possible artefacts occurring in the AC images. These artefacts, such as the arcing effects observed in the AC data [Mueller et al., 2008], display high echo strength translated by a high pixel intensity that might thus be largely represented in the ARIS Fish echograms based on pixel intensity, deceiving their reading. However, regarding the echograms based on surface area, their reading might extract debris passages in case their surface area is above the filtering threshold. This preliminary step is not designed to detect fish, only to detect targets that might be potential fish while significantly reducing the number of frames to process in the next steps.

The detection of ROI then starts by a pre-processing of the FOI before subtracting their background. Hence, the salt and pepper noise [Boncelet, 2009] presents on the acoustic videos requires a smoothing of the frame prior to their analysis [Wei et al., 2022]. Several image processing methods have been applied on acoustic videos, such as wavelets transform [Yin et al., 2020, Zang et al., 2021], Gaussian filter [French et al., 2018] or median filter [Huang et al., 2014, Jing et al., 2016]. The median filter being known to remove salt and pepper noise while preserving sharp edges [Justusson, 1981], it has been chosen as the first step of our pipeline. Despite this pre-processing not being systematically carried out on AC frames in other studies, it enables a more reliable detection of ROI with less noise. The detection of ROI corresponds to the differentiation between the foreground and the background of the AC

frames. All of the studies analysing acoustic videos based on computer vision integrate this step and different methods can therefore be found in the literature [Wei et al., 2022], such as adaptative threshold [Jing et al., 2016, dos Santos et al., 2017, French et al., 2018] or mean frame subtraction [Yin et al., 2020]. Edge detection is also used to detect ROI but requires an image manipulation to retrieve the whole ROI [Shahrestani et al., 2017, Jing et al., 2017]. Other studies use the Gaussian Mixture method of background subtraction [Kupilik and Petersen, 2014, Fernandez Garcia et al., 2023], more robust to filter noise. Among them, we choose to apply the MOG2 [Fernandez Garcia et al., 2023]. As previously mentioned, MOG2 may be optimised by tuning two parameters: the number of frames considered to model the background and the threshold beyond which pixels are not considered as foreground pixels. The settings of the two parameters of the background subtraction influence the resulting images of the moving objects, especially the degree of details of the binary images of ROI. Increasing the number of frames considered in MOG2 might improve the details of the objects but it may also increase the computation time. Besides, if only few frames precede the studied frame, the MOG2 update will hardly re-adjust as it uses sub-samples for background subtraction at each frame [Cheong et al., 2019].

Background subtraction provides binary images of the frames in which only ROI are displayed. This step of the process might highlight the fragmented target images retrieved sometimes by the AC (Figure 41). This fragmentation of a target image into several pieces, already observed by [Mueller et al., 2008], may be caused by the orientation, the position and the size of the target in the camera FOV but also depends on the AC acquisition settings. As an example, [Bothmann et al., 2016] who record larger fish at shorter window lengths than [Mueller et al., 2008] data or than our own datasets, do not encounter this issue when processing its data. Because it is a frequent limit encountered in our data, we provide an image restoration step that has never been implemented before in automatic AC data analysis to our knowledge. In the previous pipelines of AC data analysis described in the literature, the resulting binary image provided by the background subtraction is almost always exploited directly. Some studies applied a median filter to smooth them as a post-treatment [Han et al., 2009]. In addition to increasing the accuracy of the target's counts, the image restoration is also a key step for the following identification step. We have chosen to use mathematical morphological filters (a) to detect the ROI using dilation, and then (b) to restore the object's images into one single piece with successive morphological filters. We try here to restore the image as close to the truth as possible. However, the possible errors during this process may modify images erroneously, and thus lead to incorrect extractions of characteristics used for the identification. Even though we cannot really quantify them, we still need to consider this. Applying the proposed restoration process on larger datasets might however help to strengthen its performance.

6.2 Tracking target along their trajectory in AC FOV

Tracking is the second essential step required for AC analysis. The easiest tracking method is the centroid-to-centroid one that was implemented in our pipeline at first. As shown by its performance with our data, this method appears efficient when a low density of targets have distinct trajectories. It is also very sensitive to temporary lack of target's detections, a limit that has been highlighted by its performance on the BlueView data. Another tracking method presents similar limits. The alpha–beta tracking algorithm used by [Mueller et al., 2008, Kang, 2011] hardly handles high target density and target trajectories crossing each other [Wei et al., 2022]. To overcome these limits, some studies have applied the Kalman filter [Mizuno et al., 2015] or the extended Kalman filter [Huang et al., 2014, Jing et al., 2016, Jing et al., 2017] to the AC data. The Kalman filter, which predicts positions of the target from one time step to the other, allows us to overcome missing observations issues. As in the proposed pipeline, the Kalman filter gives very encouraging results. However, the number of targets' passages considered in our analysis is small, testing the method on more passages is therefore essential to strengthen the performance assessment.

The choice of the tracking method is important as it largely influences the global pipeline performance. Errors in tracking decrease the precision of the pipeline as it biases the fish count by causing multiple counting issues in addition to altering the identification. Besides, tracking is not only useful for counting, but it will also enable us to collect targets characteristics at each detection, providing information on their position, orientation, velocity and trajectory in the camera FOV. All of this information is especially useful to identify the targets nature, as well as the fish species [Bothmann et al., 2016, Mueller et al., 2008].

ROI detection and tracking are therefore the two basic first steps described in the literature [Wei et al., 2022], applied after the data conversion to detect and track targets. In the proposed pipeline a preliminary step is designed to limit the ROI detection and tracking to the FOI. As previously mentioned, during long-term monitoring, AC can record videos with no passages of objects of interest, leading to storage issues of useless data. This step thus enables us to focus only on the frames in which objects are passing through and therefore be more efficient in computation time regarding the ROI detection and tracking steps. The FOI extraction is based on the automatic reading of echograms. [Mueller et al., 2008] already highlight this idea without proposing an automatic exploitation of the echograms. The aim here was to propose an automatic reading of these echograms.

6.3 Conclusion

The first part of our pipeline is implemented with methods that have been evaluated. This pipeline can therefore be used as a semi-automated way to analyse the AC data. Hence, the

extraction of the targets by the pipeline allows an operator to visualise their passages to carry out their counting or even their species identification without visualising entirely the whole set of videos. The pipeline can also be used as part of an automatic AC data analysis. It then aims to pre-process data for the identification process while minimising potential errors that can be made at these steps. The principle of the whole method is summarised in Figure 11.

Chapter 3. Body morphology and motion: tools for species identification

Contents

A force de contempler la mer,
on finit un jour par la prendre.
Proverbe breton

1 Introduction

Anguilliform species have a characteristic serpentine shape that makes them an interesting case to study species identification on acoustic videos using the analysis of morphological characteristics.

Morphological analysis from acoustic imaging has already been used in two previous studies that aim to distinguish anguilliform from non-anguilliform targets based on computer vision methods [Mueller et al., 2008, Bothmann et al., 2016]. Extracted characteristics are common morphological information such as the target mean length or its mean surface area. Other features such as mean width or mean aspect ratio, i.e., ratio between target's width and height, are also extracted as baseline characteristics by [Bothmann et al., 2016]. Furthermore, [Mueller et al., 2008] use the compactness of the target image as a characterization of its shape. It corresponds to the ratio of the target image area to the area of a circle with the same perimeter. On their side, [Bothmann et al., 2016] designed a new metric to characterise the shape they called "watch hands". It corresponds to a scalar containing the distance between the centroid of the target image and the contour's points for different angles. In addition to these morphological characteristics, both studies also take into account motion features of the targets relying on the velocity and direction of the trajectory.

Both studies [Mueller et al., 2008, Bothmann et al., 2016] use descriptor-classifier methods. Descriptors, mainly morphological characteristics described above, aim to describe the target while classifiers aim to discriminate the classes thanks to the descriptors values. [Mueller et al., 2008] reached high performances to classify American eels (*Anguilla rostrata*) and debris, with less than 10% of misclassification using their Neural Network classifier. One of the most discriminant features for the classification was the velocity of the target as eels move faster than debris. [Bothmann et al., 2016] achieved similar conclusions for their classification between European eels (*Anguilla anguilla*), Brown trout (*Salmo trutta*) and debris, with a recall (eq. 43) of 91% and a precision (eq. 44) of 96%. Their study also highlights that including motion features improves the classification process. Regarding these studies, pairing morphological and motion analysis appears as a promising approach for multi-species identification on AC datasets.

Despite their encouraging performance, both [Mueller et al., 2008] and [Bothmann et al., 2016] pipelines do not fully fit the requirements of a generic method that my PhD work aims to reach. As mentioned in the Introduction, [Mueller et al., 2008] do not propose a fully automatic pipeline requiring successive manipulations with different software. The real-time and automatic pipeline of [Bothmann et al., 2016] shows good performance on their own data but their AC window does not exceed 6 meters, increasing the image quality. Furthermore,

they do not mention the issues raised by lower image quality such as image fragmentation [Mueller et al., 2008, Han et al., 2009]. The method performance might therefore be lower in the case of higher window length as ours. Besides, the pipeline is unfortunately not open-source and was therefore not directly deployed on our acoustic videos. Hence, as part of the surveillance of the European eel's downstream migration at the Mauzac site, long-term monitoring surveys using AC have been led by EDF. The objective is to detect migration peaks in order to adapt the turbines shutdowns. Consequently, EDF teams aim to get an almost real-time count of the European eels passages to answer management issues. However, the acoustic videos analysis by an operator is too time-consuming and therefore requires the development of a counting and automatic identification method.

To address this issue, the proposed pipeline uses some of the pre-processing methods presented in Chapter 2: extraction of FOI adapted to anguilliform species and tracking through the method of the centroid-to-centroid distance. The morphological and motion analysis is then carried out through classic morphological characteristics (Figure 49). New characteristics to define body shape as well as to estimate body length are also designed.

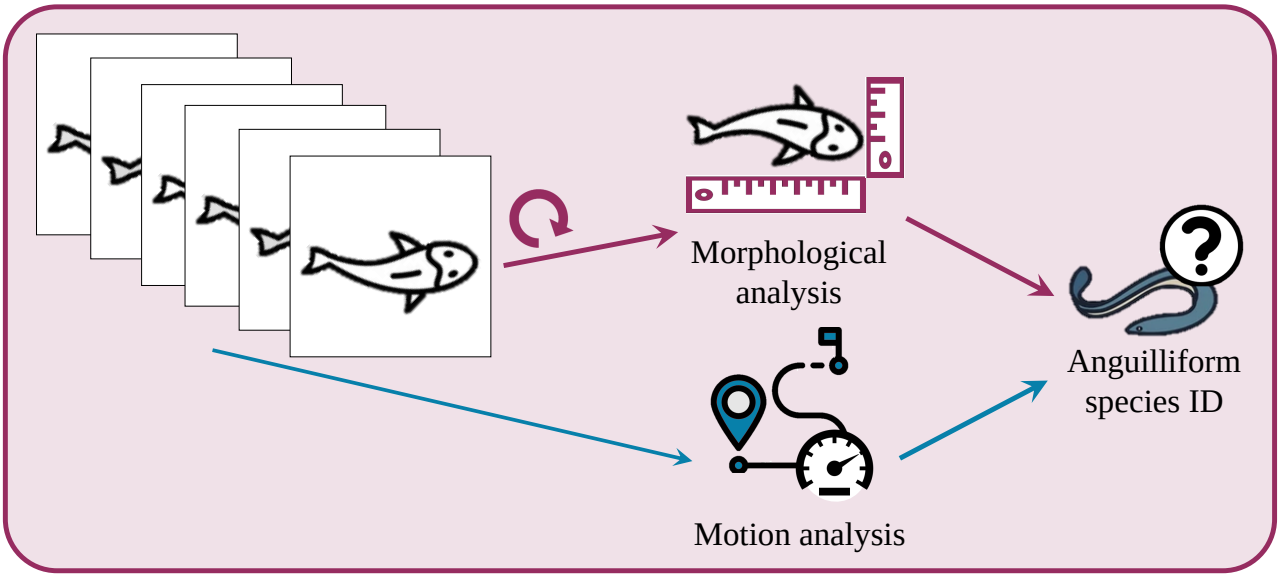


Figure 49 – Pipeline scheme of the morphological and motion analysis in order to identify anguilliform species. For each target image, the morphological analysis is carried out through length estimation and shape characterization.

Another objective of the pipeline development is to get a standardised and robust method, i.e., one single automatic method as efficient on ARIS cameras as on BlueView cameras and whose performances are similar between different monitoring sites. The pipeline has therefore been tested on the two datasets MZC-ARIS and MZC-BV as well as on the additional dataset of Port-La-Nouvelle that has a different European eel's length distribution. The pipeline and

its performances have been published in PLOS ONE [Le Quinio et al., 2023] and are the first focus of this Chapter. Then, the implementation of the pipeline on the long-term monitoring surveys of Mauzac are discussed. The information it could bring to the management of the facility regarding the European eel's downstream migration is also discussed.

Finally, to assist the operators while handling the pipeline, a user-friendly app has been developed and is presented at the end of this chapter.

2 A cross-camera morphological analysis approach: the case study of anguilliform fish

The paper published in PLOS ONE introduces the morphology and movement analysis developed for the counting and identification of anguilliform species. The morphological characteristics calculated at each target's detection are the target's length, its surface area, its global eccentricity and a metric characterising its shape that has been designed in this PhD work. However, the identification process only considers morphological characteristics measured on target's images considered as reliable, i.e. for which the calculated length is more than 75% of L_{Target}.

The motion characteristics of the target are calculated simultaneously over its entire trajectory. All of these extracted features are subjected to empirical thresholds in order to perform the identification.

The performance of this method is evaluated in regard to eels size and distance from the camera as well as to the type of acoustic camera used.

PLOS ONE

RESEARCH ARTICLE

Automatic detection, identification and counting of anguilliform fish using *in situ* acoustic camera data: Development of a cross-camera morphological analysis approach

Azénor Le Quinio[1,2]*, Eric De Oliveira[2], Alexandre Girard[3], Jean Guillard[4], Jean-Marc Roussel[1,5], Fabrice Zaoui[2], François Martignac[1,5]

1 UMR DECOD (Ecosystem Dynamics and Sustainability), Institut Agro, IFREMER, INRAE, Rennes, France, **2** EDF R&D LNHE - Laboratoire National d'Hydraulique et Environnement, Chatou, France, **3** EDF R&D PRISME - Performance, Risques Industriels et Surveillance pour la Maintenance et l'Exploitation, Chatou, France, **4** INRAE, CARRTEL, University Savoie Mont Blanc, Thonon-les-Bains, France, **5** Pole MIAME, Management of Diadromous Fish in Their Environment, OFB, INRAE, Institut Agro, University Pau & Pays Adour/E2S UPPA, Rennes, France

* azenor.le-quinio@edf.fr

OPEN ACCESS

Citation: Le Quinio A, De Oliveira E, Girard A, Guillard J, Roussel J-M, Zaoui F, et al. (2023) Automatic detection, identification and counting of anguilliform fish using *in situ* acoustic camera data: Development of a cross-camera morphological analysis approach. PLoS ONE 18(2): e0273588. https://doi.org/10.1371/journal.pone.0273588

Editor: Lorenzo Zane, University of Padova, ITALY

Received: August 9, 2022

Accepted: January 24, 2023

Published: February 24, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0273588

Data Availability Statement: Data recorded on the site of Mauzac are EDF property and are available on request at retd-lnhe-assistantes@edf.fr. The

Abstract

Acoustic cameras are increasingly used in monitoring studies of diadromous fish populations, even though analyzing them is time-consuming. In complex *in situ* contexts, anguilliform fish may be especially difficult to identify automatically using acoustic camera data because the undulation of their body frequently results in fragmented targets. Our study aimed to develop a method based on a succession of computer vision techniques, in order to automatically detect, identify and count anguilliform fish using data from multiple models of acoustic cameras. Indeed, several models of cameras, owning specific technical characteristics, are used to monitor fish populations, causing major differences in the recorded data shapes and resolutions. The method was applied to two large datasets recorded at two distinct monitoring sites with populations of European eels with different length distributions. The method yielded promising results for large eels, with more than 75% of eels automatically identified successfully using datasets from ARIS and BlueView cameras. However, only 42% of eels shorter than 60 cm were detected, with the best model performances observed for detection ranges of 4–9 m. Although improvements are required to compensate for fish-length limitations, our cross-camera method is promising for automatically detecting and counting large eels in long-term monitoring studies in complex environments.

Introduction

Active acoustic methods are widely used to study and monitor fish in marine and freshwater ecosystems [1,2]. Acoustic devices emit acoustic waves and record the echoes reflected by

authors confirm that they did not have special access privileges to the data.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

fish that pass nearby. The nature of the echoes depends on fish species and sonar specifications. Acoustic devices thus continuously record data in a non-invasive manner even at night or in turbid water, and at a long detection range (i.e. distance from the device) [2,3]. Many technical improvements have been made over time [4], among which acoustic cameras (i.e. imaging sonar) are key [5,6]. Due to their multiple beams and high frequency, these devices can project the echoes that they detect in a large volume, as 2-dimensional images. Unlike echograms produced by other types of sonar, acoustic cameras can provide high-resolution acoustic videos due to their high rate of emission-reception [5]. Therefore, fish can be visualized as they pass into the acoustic field, during which swimming behavior can be described [5,7] and morphological characteristics such as length can be measured accurately [8,9]. This provides information that is useful for identifying species [10,11]. Recently, in addition to the most commonly used models of acoustic camera–DIDSON and ARIS (Sound Metrics Corp., Bellevue, WA, USA)–new and less expensive models have become available that can scan a larger volume but have lower resolution, such as BlueView (Teledyne Technologies Inc., Thousand Oaks, CA, USA) and Oculus (Ulverston, Cumbria, United Kingdom). Based on these advantages, acoustic cameras are useful for studying long-term migration dynamics in rivers, assessing information on stock and population, and providing new insights for fish-conservation policies [6].

However, continuous recording with acoustic camera produces a large amount of data. Analysis of acoustic video by an operator is a time-consuming process that requires a degree of expertise to distinguish fish from other objects and to identify fish species. Multiple automatic or semi-automatic methods have been developed to detect and describe fish using acoustic camera datasets, listed and described in a recent review [6]. Among the studies quoted, a few authors focused on the distinction of species of interest, such as anguilliform fish, from other species [12–14]. Indeed, an operator can easily distinguish the particular body shape and swimming undulation of anguilliform fish from those of most other fish species [15–18]. Besides, the conservation status of several of anguilliform migratory species, such as the European eel (*Anguilla anguilla* [19]), American eel (*Anguilla rostrata* [20]) and Japanese eel (*Anguilla japonica* [21]), all listed on the IUCN Red List of Threatened Species, make them species of high ecological interest.

Image-processing algorithms, especially computer-vision techniques such as traditional image analysis, machine learning and deep learning, can detect objects automatically in large datasets and extract their morphological characteristics to classify them [22]. Recall and precision rates allow quantifying the ability of those algorithms to identify all target objects (recall) and distinguish them from non-target objects (precision rates). Using videos from acoustic cameras, Bothmann et al. [12] demonstrated the feasibility of using computer-vision techniques to classify European eels automatically according to their body shape and motion, highlighting promising results (recall = 91%; precision = 96%; n = 134 eels). However, the data were recorded using a DIDSON acoustic camera with a short detection range (i.e. 1–6 m), which is rarely used in long-term monitoring studies [7,17,23]. The distance of the fish from the acoustic camera, as well as its body length and orientation may decrease its detectability, because it may lead to the image of its body into distinct disconnected fragments [13,24]. Among recent studies, the deep-learning models of [14] and [25] obtained a recall of 85% when identifying American eels using ARIS sonar recorded under stable flow conditions. The authors recommended that future studies investigate the method's detection ability over long-term monitoring periods. Another recent study that used high-variability training datasets recorded under natural conditions confirmed that eels are among the most difficult fish species to identify using convolutional neural networks, a deep learning method, due to the fragmentation of eel targets in acoustic videos [26].

Based on the results and recommendations of these studies, our objective was to develop a transferrable and new method to detect anguilliform fish automatically from acoustic camera data recorded *in situ* during long-term monitoring surveys, by pairing computer-vision techniques with morphological analysis approaches to correct detection problems caused by fragmentation of fish echoes. To evaluate the effectiveness of the method and identify its potential limitations, we applied it to two datasets recorded in different rivers that had different distributions of silver eel length. The first population is mainly composed by female (body length from 50 to 100 cm [27] although males dominate the second population [28]. Body lengths of European eel male usually range from 35 to 46 cm [27]. In addition, we assessed its effectiveness in being the first cross-camera method, i.e. a method that may automatically detect and identify eels on videos recorded by two models of acoustic cameras, the ARIS and the BlueView, whose resolution and video dimension differ.

Materials and methods

Datasets

The datasets were recorded at two monitoring sites. One site (Mauzac, MZC) was located in the Dordogne River (France) at a 50-m wide inlet canal of hydropower plant (S1 Fig). The second site (Port-La-Nouvelle, PLN) was located in the 50-m wide channel between the Bages-Sigean lagoon (France) and the Mediterranean Sea [17]. At both sites, cameras were set perpendicular to the flow with the field of view (FOV) parallel to the river bottom (i.e. scanning horizontally across the channel), covering up to 10 meters from the camera at MZC and up to 14.2 meters at PLN (Table 1). The recorded images represented a top view of the water, with the X and Y dimensions corresponding to the direction of fish movement (upstream or downstream) and the detection range, respectively [5,29]. European eel populations differ between the two sites, with that at MZC containing a large proportion of large individuals (70–90 cm long, according to local fisheries and acoustic datasets) and that at PLN containing mainly small eels (30–60 cm, according to acoustic datasets). More information about the monitoring sites, their eel populations and the sonar settings can be found in [17,30,31].

Two distinct models of acoustic cameras were used to record the datasets: ARIS Explorer 1800 (at MZC and PLN) and BlueView M900-2250-130 2D (at MZC). Recording parameters

Table 1. Description of the datasets characteristics: Camera model, recording parameters and composition of the datasets, recorded at the two monitoring sites, Mauzac (MZC) and Port-La-Nouvelle (PLN).

Dataset	Development	MZC-ARIS	PLN-ARIS	MZC-BV
Monitoring site	MZC		PLN	MZC
Acoustic camera	ARIS Explorer 1800		ARIS Explorer 1800	BlueView P900-2250
Frequency (kHz)	1 800		1 800	2 250
Pixel vertical dimension (mm)	6.8		13.7	7.9
Image resolution (px)	1276×664		926×498	1238×2302
Frame rate (frames per sec.)	7		6	5
Field of view width x height (degrees)	28×14		28×14	130×20
Window length limits, from start to stop ranges (m)	0.7–9.4		1.5–14.2	0.2–10.0
Number of eels counted	24	759	788	198
Eel length (cm)	70–90	70–90	30–60	70–90
Recording duration (h)	4.0	548.5	47.5	76.0
Month recorded	Dec. 2018	Nov. 2014; Dec. 2018	Nov. 2018	Dec. 2018

https://doi.org/10.1371/journal.pone.0273588.t001

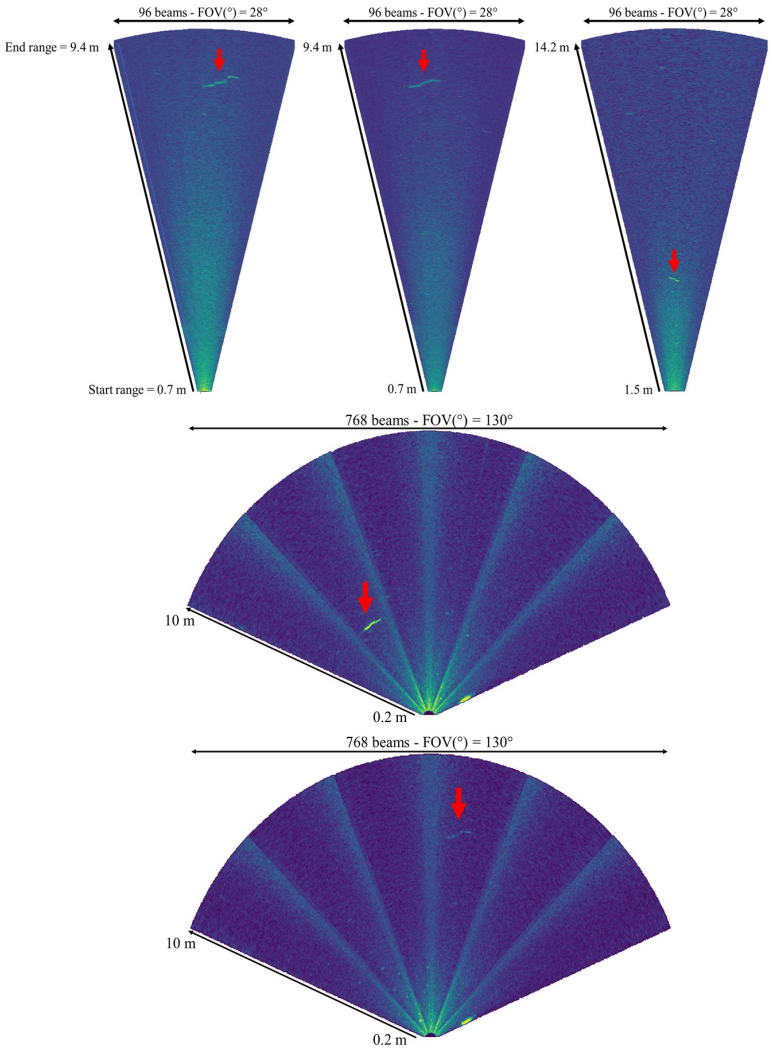


Fig 1. Example screen captures from ARIS video data recorded at (a, b) Mauzac and (c) Port-La-Nouvelle and (d, e) from BlueView video data recorded at Mauzac. Red arrows point to eels of different lengths (a: 90 cm; b: 78 cm; c: 30 cm; d: 90 cm; e: 80 cm).

https://doi.org/10.1371/journal.pone.0273588.g001

differed between cameras and sites, especially the window length and image resolution (Table 1). All data were recorded in high-frequency mode (1.80 and 2.25 MHz for ARIS and BlueView, respectively) to maximize video resolution and thus eel detection (Fig 1).

Data processing

All video datasets were watched in their entirety by experienced operators using ARIS Fish software 2.6.3 (ARIS files) or the VLC media player (BlueView files). Each European eel was counted and described by the time that it passed (both MZC and PLN), its detection range (PLN only) and its length, which was measured manually using the most representative frame

chosen by the operator (PLN only, [16]). These visual counts, usually considered the most reliable way to count fish in acoustic camera data [32,33], were used as the reference count (RC) throughout the study. The data were divided into a development dataset (4 h of ARIS data from MZC, with 24 eel passes), one ARIS evaluation dataset from each site (MZC-ARIS, 759 eel passes; PLN-ARIS, 788 eel passes) and one BlueView evaluation dataset from MZC (MZC-BV, 198 eel passes) (Table 1).

Automatic analysis pipeline

The following method has been developed using Python programming language and especially the opencv [34] and skimage [35] packages.

Our automatic analysis was decomposed in successive steps that used configuration and decision thresholds to carry out the detection, identification, and count of anguilliform fish. Those thresholds were defined empirically from the development dataset. They are adjustable and were based on four metrics: mean pixel intensity $\tilde{I}$ of the reference image (i.e. an empty image without any object passing through), pixel resolution r (mm), the frame rate fps (both imported from the recording settings) and the minimum length of the fish studied L_{eel_min} (cm), which was set by the operator before the analysis.

Step 1: Conversion to AVI. Each raw file of the datasets was converted from its proprietary format (.ARIS or.SON) into a standard video format (.AVI) using a homemade program written in Python.

Step 2: Detection of the frames of interest. Intervals of frames of interest were then extracted from the videos, i.e. the frames on which moving discontinuities were passing through the camera FOV. The discontinuities correspond to moving objects that disturb the initial background of the frames. Detection of those frames was carried out from echograms (Fig 2) based on ratios of singular values (SV, [36]) of which detailed calculation is available in S1 File. These echograms were generated by analyzing the video frame-by-frame. The aim was to reduce the method computation time by carrying out the next steps of the analysis only on these targeted frames. They were an alternative to that calculated by the ARIS Fish software, which uses the maximum intensity of echoes [37]. Our use of SV instead of maximums allows us to focus only on elongated shapes, corresponding to anguilliform ones. Hence, our tests on the development dataset videos highlight that the second and third SV explain the main part of

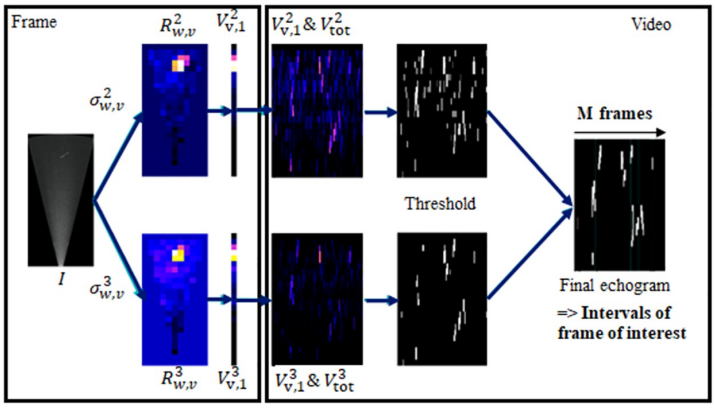


Fig 2. Workflow of the second step of the method carrying the detection of the intervals of frame of interest from the calculation of echogram based on singular values ratios.

https://doi.org/10.1371/journal.pone.0273588.g002

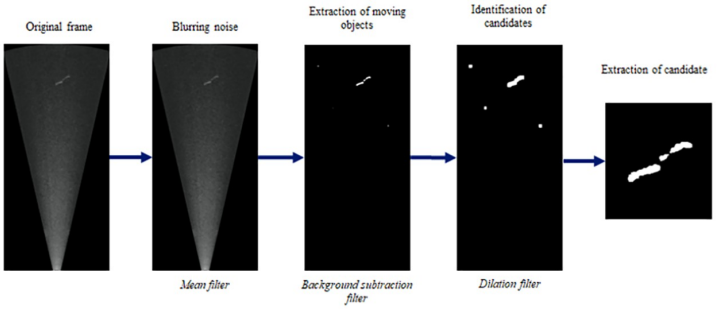


Fig 3. Workflow of the third step of the method carrying the detection of the candidates.

https://doi.org/10.1371/journal.pone.0273588.g003

the discontinuities. The third SV explains fewer discontinuities than the second one but focuses on the elongated property of the discontinuities. Analyzing the second and third ratios in a complementary way (Fig 2), thus allow us to get the precise intervals of time where elongated discontinuities were passing through and thus to target the corresponding frames.

Step 3: Detection of the candidates. Objects corresponding to the discontinuities highlighted at the previous step are extracted from the frames using a combination of image processing filters (Fig 3). A mean filter (kernel = (40 mm/*r*) x (30 mm/*r*)) followed by a Gaussian Mixture-based Background/Foreground Segmentation Algorithm (variance threshold $= 1.5 * \tilde{I}$, history = $14^{*}fps$, [38,39]) are applied to the frame in order to smooth local variations to reduce noise before isolating moving objects from the background.

A dilation (kernel = 15x15 pixels [40] is then performed on the whole resulting image. The aim is to group the objects that are close to each other making it possible to identify and isolate the regions to be studied. It will prevent an object from appearing on several regions of interest and thus prevent it from being analysed several times.

These successive treatments led to the generation of a binary image of the frame on which the different objects of interest are represented by distinct pixel regions. Objects whose main axis length was less than 25% of L_{eel_min}, i.e. global length of its body main axis which may be different from its body length in case of curved body, were removed because they were considered too short to be an eel. The rest, called "candidates", were individually extracted in a thumbnail (i.e. a square image centered on the object with a dimension of L_{eel_min} + 30 cm).

Step 4: Morphological analysis. In order to classify those candidates as eels or not, each of them was further processed by studying its physical characteristics. Its area (mm^2), orientation and overall eccentricity were calculated. The eccentricity is a measure of the non-circularity of the ellipse fitting the candidate shape and is corresponding to the ellipse focal distance divided by the length of its major axis. Its body length was estimated from its "skeleton" image, which was reconstructed for fragmented targets (see S2 File). Because the defining characteristic of anguilliform fish is their serpentine shape, with a uniform body distribution, the shape of its body was quantified by segmenting it using k-means clustering [41]. This method minimized the squared-error function based on the position of body pixels and cluster centroids. The body of the object was divided into three sections (k = 3), as a simplistic representation of the undulation of anguilliform fish with a minimum number of divisions of their body (Fig 4b and 4e). The eccentricity of each section was calculated (Fig 4c and 4f) describing the shape of the body over its entire length.

Step 5: Tracking from frame to frame. The candidates were then tracked frame-by-frame during their entire pass through the camera FOV. It was essential to avoid counting the same

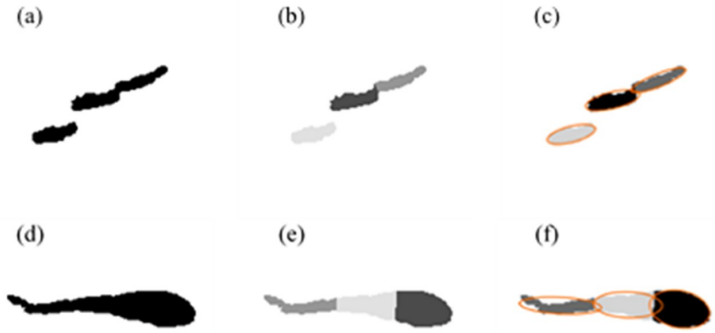


Fig 4. Steps used to calculate the shape of an echo of an eel (top) or non-eel fish (bottom) pass extracted from an ARIS camera image: (a, d) binary image, (b, e) segmentation using k-means clustering and (c, f) division of the fish body into three sections to calculate eccentricity.

https://doi.org/10.1371/journal.pone.0273588.g004

candidate more than once. Tracking also provided new and valuable information by identifying changes in the physical characteristics of the object in the successive frames in which it was detected, as well as its trajectory (S3 Fig) and velocity, among other information. To handle the tracking, the candidate's neighborhood (circle of diameter 1.5^{*} L_{eel_min}) as well as the consistency of its orientation from one detection to another ($\theta_{diff} < 25^{\circ}$), were analyzed for a short period after its last detection to determine whether an object that appeared in subsequent frames was the same candidate (S2 Fig).

Step 6: Classification process. Finally, candidates were classified as "anguilliform" or "non-anguilliform" based on characteristics calculated throughout the process (Table 2). Each candidate was first classified in each frame before considering its entire track. A candidate was classified as an anguilliform fish if it met all of the decision rules in at least five frames of its track. As for other decision rules, the minimum of five frames was set up from the development database. It corresponds to the trade-off between allowing very short passage to be considered while having enough information to carry out a reliable classification. It could be modified by the user depending on how restrictive he or she wants the classification to be.

All candidates classified as eels were listed in a.CSV output file in the directory that contained the raw dataset and correspond to the automatic counting (AC). Several characteristics

Table 2. Decision rules for characteristics used to distinguish an eel from another object as a function of the length interval of the eel population.

Scale	Characteristic	30–60 cm	60–90 cm
Frame	**Length (mm)**	$\geq 0.40 \times L_{eel\_ref}$	
	Area (mm^2)	$\geq 0.75 \times A_{thumbnail}$	
	Eccentricity	≥ 0.90	
Track	**Mean travel distance per frame (mm)**	$\geq \frac{L_{eel\_ref}}{fps}$	
	Velocity (mm/s)	Track direction was always positive along the axis of the current	
	Eccentricity	≥ 0.85	≥ 0.90
	Mean eccentricity	≥ 0.90	≥ 0.92

L_{eel_ref}, the minimum length of the fish studied, and, $A_{thumbnail}$, the area of the image centered on the object to be classified.

https://doi.org/10.1371/journal.pone.0273588.t002

were listed for each predicted anguilliform fish (each line): the video filename, mean time of passage, detection frames, detection ranges, total number of detections and the measured length of one fish per detection frame.

Evaluation of the effectiveness of the automatic eel counting method

Each video was entirely read by one experienced operator involved in data collection to compare his eels' counting to the predicted eel in the.CSV output file. Operator assessments were consequently used to calculate the number of true positives (TP, a true eel that was predicted correctly), of false positives (FP, a non-eel object incorrectly predicted as an eel) and of false negatives (FN, a true eel that was not detected/predicted). If the automatic method counted the same eel more than once, only the first count was considered a TP, while the others were considered FP, since the method also needed to accurately count the number of eels that passed into the camera FOV. Each FP was categorized to investigate the errors that occurred frequently.

The method was applied to the three evaluation datasets, and its effectiveness and the reliability of its results were assessed by calculating the recall, precision, and F1-score, as well as generating confusion matrices. Recall (Eq 1) highlights the method's ability to identify all eels that passed into the camera FOV. Precision (Eq 2) evaluates the method's ability to distinguish eels from other fish or other objects. The F1-score (Eq 3) summarized recall and precision to assess overall performance of the model. The confusion matrices (Eq 4) determined the TP, FP and FN in the three evaluation datasets. True negatives (i.e. a non-eel object that was predicted correctly) could not be counted, since the operators did not list objects other than eels when reading the videos.

$$Recall = \frac{TP}{TP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$F1\ score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{3}$$

$$Confusion\ Matrix = \begin{pmatrix} TP & FN \\ FP & - \end{pmatrix} \tag{4}$$

Linear regressions were calculated to evaluate the relationship between AC and RC and between the number of TP and RC, at an hourly resolution. The slopes of these regressions, their 95% confidence interval and their coefficient of determination (R^2) were calculated. For each pair, 70% of the data were bootstrapped 100 times to decrease bias. The slope of the linear regression of each replicate was calculated and Student's t-test was applied to the mean of the 100 slopes to determine whether it differed significantly from 1 (i.e. the slope of the 1:1 line). Differences with $p<0.05$ were regarded as statistically significant. The two regressions were compared to illustrate the benefit of having an operator validate each prediction to assess the accuracy of the eel count.

Finally, factors that may have influenced the method were evaluated to assess its genericity (i.e. ability to operate well, regardless of the monitoring site and the acoustic camera model). The influence of three factors on the recall, precision and F1-scores were qualitatively studied. The first was the acoustic camera model, for which we analyzed the confusion matrices and

Table 3. Confusion matrices obtained for the Mauzac site/ARIS camera (MZC-ARIS), Mauzac site/BlueView camera (MZC-BV) and Port-La-Nouvelle site/ARIS camera (PLN-ARIS) datasets.

	MZC				PLN			
ARIS	**MZC-ARIS**	*Pred. Eels*	*Pred. Others*	*Total*	**PLN-ARIS**	*Pred. Eels*	*Pred. Others*	*Total*
	Eels	556	197	753	*Eels*	349	439	788
	Others	110	*	-	*Others*	485	*	-
	Total	666	-	-	*Total*	832	-	-
BV	**MZC-BV**	*Pred. Eels*	*Pred. Others*	*Total*				
	Eels	139	59	198				
	Others	48	*	-				
	Total	187	-	-				

Lines are true eels and true other species fish, columns are eels and other species fish predicted by the method * True negatives were not counted.

https://doi.org/10.1371/journal.pone.0273588.t003

metrics for a 73-hour period common to the MZC-ARIS and MZC-BV datasets (RC = 56 and 174, respectively). The second factor was the eel's detection range, based on TP and FN distributions, and the change in recall and precision as detection range increased for PLN-ARIS. The third factor was the eel's measured length, based on the relationship between the recall and eel length for PLN-ARIS.

Results

Performance of the method

Confusion matrices showed differences between predicted eel detection and the eels observed by operators for the three evaluation datasets (Table 3). MZC datasets had higher F1-scores (72–78%) than the PLN dataset (43%) (Table 4). ARIS and BlueView had similar recall (74% and 70%, respectively), but ARIS had higher precision (84%, vs. 74% for BlueView).

Regression slopes between RC and AC differed significantly ($p < 0.001$) from 1 for all three datasets, but large positive linear association are observed for MZC-ARIS ($R^2 = 0.93$) and MZC-BV ($R^2 = 0.95$), and weaker for PLN-ARIS ($R^2 = 0.41$), with a slope of 0.47 (Fig 5a–5c). Regression slopes between the number of TP and RC also differed significantly ($p < 0.001$) from 1 for all three datasets, and the correlations for all datasets were stronger than those between RC and AC but differed more between monitoring sites (Fig 5d–5f). The correlations were strong for PLN ($R^2 = 0.79$), which highlighted a large number of FP at this site and were slightly stronger for MZC-ARIS and MZC-BV ($R^2 = 0.97$ and 0.96, respectively), with similar slopes (0.72) and a small 95% confidence interval (0.01 and 0.02, respectively).

One source of errors for MZC-ARIS and PLN-ARIS (37% and 75% of the FP, respectively) (Table 5) was misidentification due to confusion with other fish species (S4c Fig) or debris. It was the second-largest source of errors (42%) for MZC-BV, while the largest source was tracking errors that resulted in counting the same eel more than once (50%). Other sources of errors

Table 4. Performance metrics for the Mauzac site/ARIS camera (MZC-ARIS), Mauzac site/BlueView camera (MZC-BV) and Port-La-Nouvelle site/ARIS camera (PLN-ARIS) datasets.

Metric	MZC-ARIS	MZC-BV	PLN-ARIS
Recall	73.8%	70.2%	44.3%
Precision	83.5%	74.3%	41.9%
F1-score	78.3%	72.2%	43.1%

https://doi.org/10.1371/journal.pone.0273588.t004

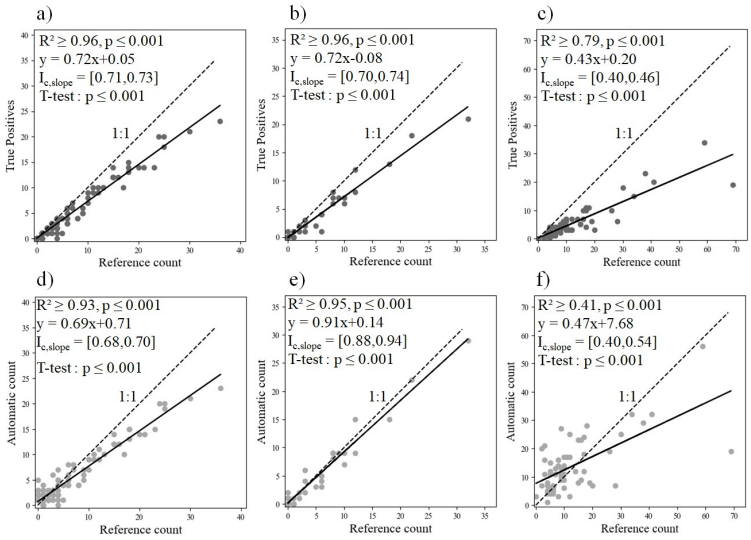


Fig 5. The method's (a, b, c) true positives and (d, e, f) automatic count of eels as a function of the reference count for (a, d) Mauzac site/ARIS camera (MZC-ARIS), (b, e) Mauzac site/BlueView camera and (c, f) Port-La-Nouvelle site/ARIS camera. Each plot represents the number of eels counted in one hour (MZC-ARIS: n = 199; MZC-BV: n = 55; PLN-ARIS: n = 60). $I_{c,slope}$ is the 95% confidence interval of the slope.

https://doi.org/10.1371/journal.pone.0273588.g005

were arcing effects (S4a Fig), artefacts of high intensity resulting of returning echoes, which deform fish by giving them a more streamlined shape [13], which was observed in MZC-ARIS and PLN-ARIS (36% and 5%, respectively), but not in MZC-BV. Additionally, small fish swimming near each other (S4b Fig) were sometimes detected as one larger individual (20% and 8% of the FP in MZC-ARIS and PLN-ARIS, respectively). Because tracking errors were not considered errors when identifying species, a modified precision (i.e. identification precision) was calculated that excluded them from the FP. Identification precision on MZC-BV was similar to the precision for the ARIS datasets, but was much larger than the precision for identification and counting for MZC-BV (85% for identification and 74% for identification and counting).

Analysis of factors that influenced performances of the method

The two acoustic camera models at MZC had similar recall and precision (Table 6). For PLN-ARIS, performances peaked 4–9 m from the camera, for which the F1-score exceeded

Table 5. Percentage of false positives (FP) in each error category observed, total precision and identification precision (i.e. excluding tracking errors) for each dataset: Mauzac site/ARIS camera (MZC-ARIS), Mauzac site/BlueView camera (MZC-BV) and Port-La-Nouvelle site/ARIS camera (PLN-ARIS). Bold text indicates the largest percentage per dataset.

		MZC-ARIS	MZC-BV	PLN-ARIS
Error category	Arcing effect	36.3% (n = 40)	0% (n = 0)	4.9% (n = 24)
	Merged fish	20.0% (n = 22)	8.3% (n = 4)	3.1% (n = 15)
	Other type	**37,3%** (n = 41)	41.7% (n = 20)	**75.3%** (n = 365)
	Tracking	6.4% (n = 7)	**50%** (n = 24)	16.7% (n = 81)
Total precision		83.4%	74.3%	41.9%
Identification precision		84.4%	85.3%	44.3%

https://doi.org/10.1371/journal.pone.0273588.t005

Table 6. Confusion matrices and performance metrics obtained for a common 73-hour period at Mauzac recorded by ARIS and BlueView cameras.

ARIS	Pred. Eels	Pred. Others	Total	BlueView	Pred. Eels	Pred. Others	Total
Eels	38	18	56	Eels	122	52	174
Others	13	*	-	Others	46	*	-
Total	51	-	-	Total	168	-	-
		Recall	67.9%			Recall	70.1%
		Precision	74.5%			Precision	72.6%
		F1-score	64.8%			F1-score	68.9%

* True negatives were not counted.

https://doi.org/10.1371/journal.pone.0273588.t006

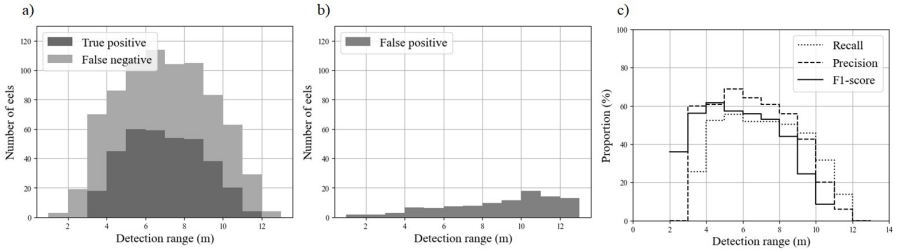


Fig 6. Performance metrics of the method as a function of the detection range (m) for the PLN-ARIS dataset: a) cumulative true positive and false negative distributions and recall, b) false positive distribution and precision, and c) F1-scores.

https://doi.org/10.1371/journal.pone.0273588.g006

40% (Fig 6a). Beyond 9 m, the precision progressively decreased, with more than 50% of the FP recorded beyond this distance (Fig 6b). Similarly, only 35% of eels that passed more than 10 m and 20% of those that passed within 4 m from the camera were detected (Fig 6c), which resulted in a decrease in recall for these ranges. Eel length also influenced the method's performance. Despite a recall of 44.3% for PLN-ARIS, eel length distributions from TP records and RC were similar (Fig 7). The method's recall decreased below a threshold length of 37 cm but remained near 60% above it. Only one eel larger than the targeted range of 30–60 cm was detected at PLN (78 cm long).

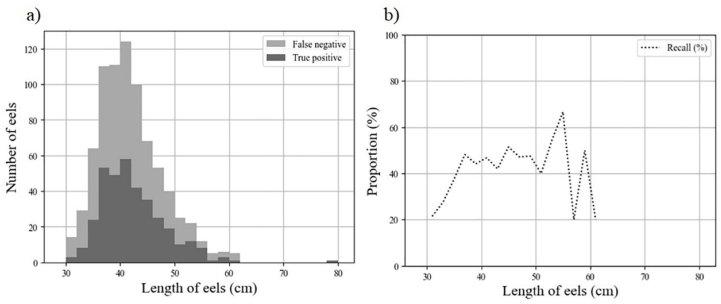


Fig 7. Performance metrics (cumulative false negative and true positive distributions and recall values) of the method as a function of the measured length of eels (cm) for the PLN-ARIS dataset.

https://doi.org/10.1371/journal.pone.0273588.g007

Discussion

The method developed to identify anguilliform fish using acoustic camera records performed well compared to those in previous studies [12,14]. The method was able to automatically identify 74% of the large European eels that passed into the ARIS sonar field of view (recall) with 84% precision (i.e. 16% FP). For the same species, Bothmann et al. [12] obtained a higher performance (recall = 91%, precision = 96%) with a DIDSON camera but on a 5m long detection range, *i.e.* 2–3 times shorter than our study. With better image resolution, the fragmented targets were likely not a serious problem for their dataset. Zang et al. [14] also obtained a similar good performance (recall = 84%) on a few individuals of American eels, with no other fish species in their dataset. Moreover, both studies [12,14] used data from a single acoustic camera model at a single monitoring site. The method developed in the present study was tested with two common and different models of acoustic camera and aims to be transferable to other models of acoustic camera. Following the pioneer DIDSON and ARIS cameras, other high-frequency multibeam sonars are now available, notably BlueView and Oculus [42]. These devices are less expensive and compensate their lower resolution by having a wider FOV: they thus cover more than four times as much volume as DIDSON or ARIS. When ARIS was compared to BlueView at MZC, the latter counted three times as many eels during the same recording period. Despite BlueView's lower resolution, the method developed obtained similar recall and precision with it as with ARIS, which highlights the method's genericity. In our study site where both camera types were set up, the BlueView wider FOV increased the capacity of the camera to detect large European eels despite its poorer resolution. Furthermore, our method of counting and detection performed similarly as on ARIS camera for the same population of eel. Previous studies have already shown the potential of the BlueView camera in similar settings [42].

However, more than 50% of the FP recorded in MZC-BV resulted from multiple counting of eels. Considering only the method's ability to distinguish eels from other species or objects, and not the counting accuracy, the precision for MZC-BV was 85.3%, which was similar to that of MZC-ARIS (84.4%). This highlights the method's strong ability to distinguish eels from other objects using either camera model. Because the method also aims to count eels accurately, multiple counts are considered an error, and they decreased the precision for BV-MZC slightly to 74.3%. The larger number of multiple counts in BlueView videos than in ARIS videos was due to the greater difficulty in accurately and continuously tracking the same eel along its path. This error in BlueView videos could have come from two sources: (a) structural noise that appears every 21˚, which prevents capturing an accurate image of objects in the impacted frame quadrants, and (b) the much larger field of view, which makes it more likely to lose an object as it moves and to find it again a little further in the FOV. Post-processing the paths using post-tracking reconstruction to exclude the noise quadrants could help remove duplicates from the eel count.

Despite these frequent errors, the method underestimates the eel count to a similar degree in MZC-ARIS and MZC-BV, with a regression slope of 0.72, a small 95% confidence interval and a strong correlation, despite the differences in the beam formation process of both cameras and in the consecutive resolutions of their images. Those correlations between the number of eels counted by an operator and the eels automatically detected highlight that a constant percentage of eels is efficiently identified by our method on both datasets on the datasets recorded by both camera models even if slopes are significantly different from the 1:1 line. This information could help future studies determine a correction factor site-dependent for the method's count to reach a similar efficiency than an entire reading by an operator.

During their downstream migration, European silver eels display a wide range of body lengths, as well as sexual dimorphism (i.e. males are shorter than females [31]). Overall, sex determination in the European eel is partly related to the local environmental conditions they encounter during their freshwater phase, such as eel density, recruitment and watershed characteristics [43,44]. Our method was more effective for sites dominated by females (recall = 74%, precision = 83%) and performed moderately well when the eels were shorter than 60 cm (recall = 44%, precision = 42%). This limitation was likely due to the camera resolution and fish length. Information about morphology is severely limited when an anguilliform fish is characterized by only a few pixels. Fernandez Garcia et al. [26] obtained similar results with conventional neural networks: their results showed that length had a clear influence on eel recall, decreasing from 65% for eels longer than 60 cm to 22% for those shorter than 60 cm. Despite length-biased performance, we found that the predicted length distribution was consistent with that observed by the operators (Fig 6).

In acoustic camera videos, the horizontal dimension of pixels, or cross-range resolution [6] depends on the detection range, which influences an operator's visualization of fish [45] and may decrease measurement accuracy [46,47]. Our results for the PLN-ARIS dataset highlight the method's robust performances for eels 4–9 m from the camera. Eels closer to the camera were not detected well since their echoes were sometimes hidden or corrupted by arcing effects. These errors are a known issue in DIDSON videos due to fish or highly reflective targets moving near the camera [13,46], but they were not observed in BlueView videos. Mitigating the signal intensity on the recorded videos may decrease the influence of those arcing effects but may also affect the efficiency of automatic fish detections. The specific shape of these artifacts (i.e. the arc of a circle) could be addressed by specific pre-processing of each frame. The ability to identify all eels that passed in the camera FOV peaked 4–10 m from the camera, with a recall of ca. 60%. Similarly, the precision (i.e. the ability to distinguish eels from other objects) also depended on the detection range. Precision peaked 4–9 m from the camera, with a precision of 60%. The number of FP increased greatly beyond 9 m, with errors due mainly to misidentifying other species or debris as eels. This was likely because the increase in the pixel dimension prevented effective identification of small eels. Future studies should focus on whether our model is more effective for detecting and identifying larger eels beyond 9 m.

Unlike the frame resolution of the DIDSON camera, that of ARIS can be set by the operator by adjusting the number of samples along the vertical axis. The vertical resolution can thus range from 3–19 mm, which causes differences in width for fish that move perpendicular to the sonar FOV. Theoretically, we assume that fish identification will improve at a higher resolution. The difference in resolution between MZC and PLN may also have influenced performances at PLN, making it more difficult to distinguish eels from other objects, especially those further from the camera (63% of FP in the PLN dataset). An overly low resolution may also increase the fragmentation of eel echoes that cause discontinuities and the need to rebuild the eel skeleton. In addition to camera recording parameters, future studies should focus on environmental conditions during recording and their influence on identification. Among other factors, water flow can influence the migration activity of silver eels [48] as well as influencing image quality via the number and speed of objects that pass in the camera FOV. Thus, it could have influenced the method's performance by increasing the number of objects detected and shortening the interval of frames of each pass.

Conclusion

Our study demonstrates the feasibility of using computer vision and morphological analysis to automatically identify large anguilliform fish *in situ*, using two acoustic camera models with

different characteristics, up to a range of 10 m, even when images of fish become fragmented. Unless fish length emerges as a limitation of this method–and future studies should be performed to define its operational boundaries more clearly. Our innovative method can provide relevant ecological data from acoustic cameras during long-term monitoring, in a faster way compared to reading of the entire dataset by a human operator, pushing further the advantage of using acoustic cameras to monitor migratory fish populations without any disturbance on their behavior and health conditions face to more intrusive methods as stow nets.

The results reveal that the automatic method with no post-processing could be useful for monitoring migration dynamics of a large anguilliform fish population. Although it tends to underestimate the total number of fish, it can identify their migration peaks and serve as a proxy of the total number of fish passing through the camera FOV. In a semi-automated application, an experienced operator could validate each predicted anguilliform fish, which would increase the accuracy of the fish count and avoid FP detections. Having a proxy of eel passes is especially useful for the operation of hydropower plants. Future development could include a user-friendly application to monitor *in situ* eels in real time, which could encourage more precise management of turbine shutdowns that optimizes production costs while hurting fewer eels. Besides, the problem of large acoustic data storage can be alleviated by running the analysis right after recording and before archiving the data; the use of higher resolution settings should be much easier too.

Although the method was successful for large silver eels, remaining issues could be investigated to improve it further. Machine learning could be applied to improve the accuracy of the empirically defined thresholds used in the morphological analysis. Fish length, detection range and resolution influence the method's performance. Better understanding the influence of each factor as well as their interactions is a potential area for improvement. Currently, caution is required when using our approach since the automatic detection of small (< 60 cm long) eels does not perform very well and counting might be undervalued. Finally, the method was designed and tested only for European eels, but it could be tested for other species with similar undulation and a serpentine body shape, such as the American eel and sea lamprey (*Petromyzon marinus*). Morphological analysis of moving fish is a promising and timesaving approach for using acoustic cameras to identify and count other fish species.

Supporting information

S1 Fig. Figure of the installation of the two acoustic cameras at MZC site.
(TIF)

S2 Fig. Workflow of the reconstruction process of skeleton in case of fragmented object. a) Binary image of the candidate's body after background subtraction; b) Binary image of the candidate skeleton; c) Binary image of the candidate reconstructed skeleton.
(TIF)

S3 Fig. Tracking conditions between two detections of an eel at n frames interval.
(TIF)

S4 Fig. Successive positions and shape of an eel along its movement in the camera beam.
(TIF)

S5 Fig. Examples of three of the main categories of errors made by the method. a) Example of an arcing effect; b) Example of the method misidentification due to fish swimming closely to each other, c) Example of the method misidentification as an eel of another fish species.
(TIF)

S1 File. Calculation of the echograms of singular values ratios.
(PDF)

S2 File. Reconstruction and extraction of skeleton characteristics.
(PDF)

Acknowledgments

This study was supported by Electricité de France and the French Association Nationale de la Recherche et de la Technologie. We sincerely thank the CEFREM team (University of Perpignan) for the PLN dataset and its visual analysis. We are grateful to the EDF and MIGADO teams who visually analyzed the Mauzac datasets. Finally, we sincerely thank Guy d'Urso for his help and wise advices.

Author Contributions

Conceptualization: Azénor Le Quinio, Eric De Oliveira, François Martignac.

Data curation: Azénor Le Quinio, Eric De Oliveira, Alexandre Girard, Fabrice Zaoui.

Formal analysis: Azénor Le Quinio, Alexandre Girard, Fabrice Zaoui.

Funding acquisition: Eric De Oliveira.

Investigation: Azénor Le Quinio.

Methodology: Azénor Le Quinio, Alexandre Girard, Fabrice Zaoui.

Project administration: Eric De Oliveira, Jean Guillard, Jean-Marc Roussel, François Martignac.

Software: Azénor Le Quinio, Alexandre Girard, Fabrice Zaoui.

Supervision: Eric De Oliveira, Jean Guillard, Jean-Marc Roussel.

Validation: Eric De Oliveira, Jean-Marc Roussel, François Martignac.

Visualization: Azénor Le Quinio.

Writing – original draft: Azénor Le Quinio, Eric De Oliveira, François Martignac.

Writing – review & editing: Azénor Le Quinio, Eric De Oliveira, Alexandre Girard, Jean Guillard, Jean-Marc Roussel, Fabrice Zaoui, François Martignac.

References

1. Rudstam LG, Jech JM, Parker-Stetter SL, Horne JK, Sullivan PJ, Mason DM. Fisheries Acoustics. In: Fisheries Techniques, Third edition. Zale A.V., Parrish D.L., Sutton T.M. 2012. p. 40.
2. Simmonds J, MacLennan DN. Fisheries Acoustics: Theory and Practice. 2nd edition. Oxford: Blackwell Science; 2005. (Fisheries Series).
3. Foote K. Acoustic Methods: Brief Review and Prospects for Advancing Fisheries Research. In 2009. p. 313–43.
4. Trenkel V, Ressler P, Jech M, Giannoulaki M, Taylor C. Underwater acoustics for ecosystem-based management: State of the science and proposals for ecosystem indicators. Marine Ecology Progress Series. 2011; 442.
5. Martignac F, Daroux A, Bagliniere JL, Ombredane D, Guillard J. The use of acoustic cameras in shallow waters: new hydroacoustic tools for monitoring migratory fish population. A review of DIDSON technology. Fish and Fisheries. 2015; 16(3):486–510.
6. Wei Y, Duan Y, An D. Monitoring fish using imaging sonar: Capacity, challenges and future perspective. Fish and Fisheries [Internet]. 2022; n/a(n/a). Available from: https://doi.org/10.1111/faf.12693

7. Burwen D, Fleischman S, Miller J. Accuracy and Precision of Salmon Length Estimates Taken from DIDSON Sonar Images. Transactions of the American Fisheries Society. 2010; 139:1306–14.
8. Daroux A, Martignac F, Nevoux M, Baglinière JL, Ombredane D, Guillard J. Manual fish length measurement accuracy for adult river fish using an acoustic camera (DIDSON). Journal of Fish Biology. 2019; 95. https://doi.org/10.1111/jfb.13996 PMID: 31049959
9. Helminen J, Dauphin G, Linnansaari T. Length measurement accuracy of Adaptive Resolution Imaging Sonar (ARIS) and a predictive model to assess adult Atlantic salmon (*Salmo salar*) into two size categories with long-range data in a river. Journal of Fish Biology. 2020; 97.
10. Able K, Grothues T, Rackovan J, Buderman F. Application of Mobile Dual-frequency Identification Sonar (DIDSON) to Fish in Estuarine Habitats. Northeastern Naturalist. 2014; 21:192–209.
11. Jones R, Griffin R, Unsworth R. Adaptive Resolution Imaging Sonar (ARIS) as a tool for marine fish identification. Fisheries Research. 2021; 243:106092.
12. Bothmann L, Windmann M, Kauermann G. Realtime classification of fish in underwater sonar videos. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2016; 65:n/a–n/a.
13. Mueller AM, Mulligan T, Withler P. Classifying Sonar Images: Can a Computer-Driven Process Identify Eels? North American Journal of Fisheries Management—NORTH AM J FISH MANAGE. 2008; 28:1876–86.
14. Zang X, Yin T, Hou Z, Mueller RP, Deng ZD, Jacobson PT. Deep Learning for Automated Detection and Identification of Migrating American Eel *Anguilla rostrata* from Imaging Sonar Data. Remote Sensing [Internet]. 2021; 13(14). Available from: https://www.mdpi.com/2072-4292/13/14/2671.
15. Keeken O, Hal R, Winter HV, Tulp I, Griffioen A. Behavioural responses of eel (*Anguilla anguilla*) approaching a large pumping station with trash rack using an acoustic camera (DIDSON). Fisheries Management and Ecology. 2020; 27.
16. Lagarde R, Peyre J, Amilhat E, Mercader M, Prellwitz F, Gael S, et al. In situ evaluation of European eel counts and length estimates accuracy from an acoustic camera (ARIS). Knowledge and Management of Aquatic Ecosystems. 2020; 421:44.
17. Lagarde R, Peyre J, Amilhat E, Bourrin F, Prellwitz F, Gael S, et al. Movements of Non-Migrant European Eels in an Urbanised Channel Linking a Mediterranean Lagoon to the Sea. Water. 2021; 13:839.
18. Lenihan ES, McCarthy TK, Lawton C. Use of an acoustic camera to monitor seaward migrating silver-phase eels (*Anguilla anguilla*) in a regulated river. Ecohydrology & Hydrobiology. 2019; 19(2):289–95.
19. Dekker W, Beaulaton L. Climbing back up what slippery slope? Dynamics of the European eel stock and its management in historical perspective. ICES J Mar Sci. 2015; 73(1):5–13.
20. Macgregor R, Casselman JM, Allen WA, Haxton T, Dettmers JM, Mathers A, et al. Natural heritage, anthropogenic impacts, and biological issues related to the status and sustainable management of American eel: a retrospective analysis and management perspective at the population level. 2009; 69:713–40.
21. Kaifu K, Yokouchi K. Increasing or decreasing?—current status of the Japanese eel stock. Fisheries Research. 2019.
22. LeCun Y, Bengio Y, Hinton G. Deep Learning. Nature. 2015; 521:436–44. https://doi.org/10.1038/nature14539 PMID: 26017442
23. Helminen J, Linnansaari T. Object and behavior differentiation for improved automated counts of migrating river fish using imaging sonar data. Fisheries Research. 2021; 237:105883.
24. Han J, Honda N, Asada A, Shibata K. Automated acoustic method for counting and sizing farmed fish during transfer using DIDSON. Fisheries Science. 2009; 75:1359–67.
25. Yin T, Zang X, Hou Z, Jacobson P, Mueller R, Deng Z. Bridging the Gap between Laboratory and Field Experiments in American Eel Detection Using Transfer Learning and Convolutional Neural Network. In 2020.
26. Fernandez Garcia G, Martignac F, Nevoux M, Beaulaton L, Corpetti T. A deep neural network for multi-species fish detection using multiple acoustic cameras. 2021 Sep; Available from: https://doi.org/10.48550/arXiv.2109.10664
27. Durif C, Dufour S, Elie P. The silvering process of the eel: a new classification from the yellow resident stage to the silver migrating stage. Journal of Fish Biology—J FISH BIOL. 2005 Apr; 66:1025–43.
28. Amilhat E, Farrugio H, Lecomte-Finiger R, Gael S, Sasal P. Silver eel population size and escapement in a Mediterranean lagoon: Bages-Sigean, France. Knowledge and Management of Aquatic Ecosystems (1961–9502) (EDP Sciences), 2008–09, N 390–391, P 2–11. 2008 Jan;5.
29. Hughes J, Hightower J. Combining Split-Beam and Dual-Frequency Identification Sonars to Estimate Abundance of Anadromous Fishes in the Roanoke River, North Carolina. North American Journal of Fisheries Management. 2015; 35:229–40.

30. ECOGEA. Evaluation de l'efficacité du dispositif de dévalaison au niveau de l'aménagement hydroélectrique EDF de Tuilières (Dordogne) pour les anguilles d'avalaison. Résultats des tests d'automne-hiver 2012–2013. ECOGEA; 2014. Report No.: E121015.

31. Tesch FW. The eel. Blackwell. London; 2003. 397 p.

32. Holmes J, Cronkite G, Enzenhofer H, Mulligan T. Accuracy and precision of fish-count data from a "dual-frequency identification sonar" (DIDSON) imaging system. Ices Journal of Marine Science—ICES J MAR SCI. 2006; 63:543–55.

33. Maxwell S, Gove N. Assessing a dual-frequency identification sonars' fish-counting accuracy, precision, and turbid river range capability. The Journal of the Acoustical Society of America. 2008; 122:3364–77.

34. Bradski G. The OpenCV Library. Dr Dobb's Journal of Software Tools. 2000.

35. Van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, et al. scikit-image: image processing in Python. PeerJ. 2014; 2:e453. https://doi.org/10.7717/peerj.453 PMID: 25024921

36. Golub G, Reinsch C. Singular value decomposition and least squares solution. Numerische Mathematik. 1969; 14:42.

37. Sound Metrics Corp. ARISFish Software User Guide: Echogram Counting Tutorial. Sound Metrics Corp; 2014.

38. Zivkovic Z. Improved Adaptive Gaussian Mixture Model for Background Subtraction. In: Proceedings—International Conference on Pattern Recognition. 2004. p. 28–31 Vol. 2.

39. Zivkovic Z, Van der Heijden F. Efficient adaptive density estimation per image pixel for the task of background subtraction. Pattern Recognition Letters. 2006; 27:773–80.

40. Gonzalez RC, Woods RE. Digital Image Processing (3rd Edition). USA: Prentice-Hall, Inc.; 2006.

41. Lloyd SP. Least squares quantization in PCM. IEEE Trans Inf Theory. 1982; 28:129–36.

42. Capoccioni F, Leone C, Pulcini D, Cecchetti M, Rossi A, Ciccotti E. Fish movements and schooling behavior across the tidal channel in a Mediterranean coastal lagoon: An automated approach using acoustic imaging. Fisheries Research. 2019; 219.

43. Parsons J, Vickers K, Warden Y. Relationship between elver recruitment and changes in the sex ratio of silver eels *Anguilla anguilla* L. migrating from Lough Neagh, Northern Ireland. Journal of Fish Biology. 1977; 10:211–29.

44. Vollestad LA, Jonsson B. A 13-Year Study of the Population Dynamics and Growth of the European Eel Anguilla anguilla in a Norwegian River: Evidence for Density-Dependent Mortality, and Development of a Model for Predicting Yield. Journal of Animal Ecology. 1988; 57(3):983–97.

45. Burwen D, Fleischman S, Miller J. Evaluation of a dual-frequency imaging sonar for detecting and estimating the size of migrating salmon. Alaska Department of Fish and Game, Division of Sport Fish, Research and Technical Service. 2007.

46. Burwen D, Fleischman S, Miller J. Accuracy and Precision of Salmon Length Estimates Taken from DIDSON Sonar Images. Transactions of the American Fisheries Society. 2010; 139:1306–14.

47. Tušer M, Frouzova J, Balk H, Muška M, Mrkvicka T, Kubecka J. Evaluation of potential bias in observing fish with a DIDSON acoustic camera. Fisheries Research. 2014; 155:114–21.

48. Reckordt M, Ubl C, Wagner C, Frankowski J, Dorow M. Downstream migration dynamics of female and male silver eels (Anguilla anguilla L.) in the regulated German lowland Warnow River. Ecology of Freshwater Fish. 2014; 23(1):7–20.

3 Operational case study: the European eels of the Mauzac hydropower plant

Being the first obstacle for the European eels during their downstream migration, the Mauzac hydropower plant is a site that has to be monitored. Hence, a regulatory obligation imposes turbine shutdowns during their downstream migration period. In order to optimise these shutdowns, modelling studies aiming to predict the dynamics of downstream movements have been launched [E. De Oliveira, personal communication], integrating results from radio telemetry or scientific fishery studies [ECOGEA, 2014]. Since 2014, the Mauzac site has been equipped with AC to meet multiple objectives:

- Detect the beginning and the end of the European eel's migration with *in situ* observations.
- Monitor the downstream migration over long periods of time and for several years.
- Validate and optimise the prediction model that, for now, plans the turbines' shutdowns.
- Assess the AC performances to estimate the downstream European eel's stock.

As mentioned in Chapter 1, a scientific fishery campaign has been carried out at Mauzac. For the last 14 days of the fisheries in November 2014, the nets have been installed at the entrance of the inlet canal, upstream of the water section covered by the AC, in order to estimate the AC ability to record eels passages. The eels counted by an operator in acoustic videos are compared to fisheries results on a 14 nights period. In a similar period, AC data were able to record 10 times more eels passages than eels captured by the fisheries (Figure 50), confirming the high potential of AC.

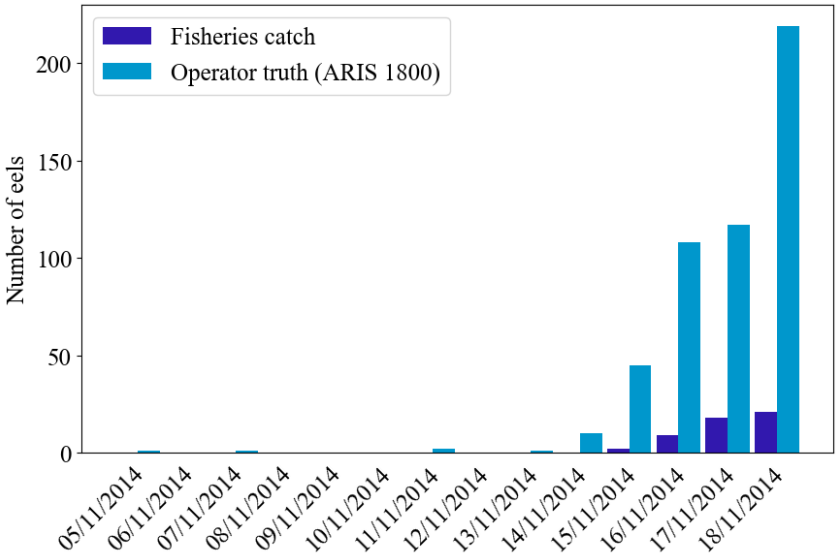


Figure 50 – Comparison of the number of European eels counted from the fisheries and from the ARIS 1800 videos reviewed by an experienced operator, on 14 nights of November 2014.

3.1 Reliability of automatic detection of European eel's migration

The ability to correctly identify the migration periods of European eels from automatic counting is first evaluated. As introduced in Chapter 1, the ARIS 1800 camera has recorded during three migration periods. Over 21 days from 2014-2015 and 2018, 783 eels passages have been identified by the operators. The automatic method is applied on the same videos allowing to obtain the number of True Positives (TP), i.e., true eels that have been automatically counted as such, False Positives (FP), i.e., non-eel objects that have been automatically counted as eels, and False Negatives (FN), i.e., eels that have not been counted by the automatic method. The distribution of counts provided by the automatic method is compared to the operator truth as recall (eq. 43) and precision (eq. 44) values.

$$Recall = \frac{TP}{TP + FN} \quad \text{(eq. 43)}$$

$$Precision = \frac{TP}{TP + FP} \quad \text{(eq. 44)}$$

With, TP, the number of true positives; FN, the number of false negatives; and, FP, the number of false positives.

From both monitoring periods of 2014 and 2018 (Figure 51.a and b), when the number of eels daily passages based on the operator truth is superior to 20 (16/11/2014 to 19/11/2014 and 04/12/2018 to 05/12/2018), the automatic analysis underestimates the number of eels passages. Conversely, when the number of eels based on the operator truth is lower, i.e., inferior to 20 daily eels passages, the automatic analysis overestimates the number of passages. However, results of December 2018 also show that in some cases, the number of errors of the method (FP) compensates the missing eels, reaching a similar total count of the method with the operator truth as it can be observed on the 05/12/2018, reaching a 100% recall (Figure 51).

Hence, recall rate allows to determine the performance of the method to correctly count all eels passages. It reaches encouraging results on the last days of November 2014 and on the early days of December 2018 (61.5 % and 74.1% on 04/12/2018 and 05/12/2018) when the number of eels passages is considered high (Figure 51.a and b). However, a good method performance is a balance between maximisation of recall and precision rates. Precision rate allows to determine the ability of the method to differentiate European eels from other objects. High precision is observed on the last days of November 2014 and early days of December 2018 (84.2 % and 74.1% on 04/12/2018 and 05/12/2018), which paired with the good recall rate gives encouraging performance of the method.

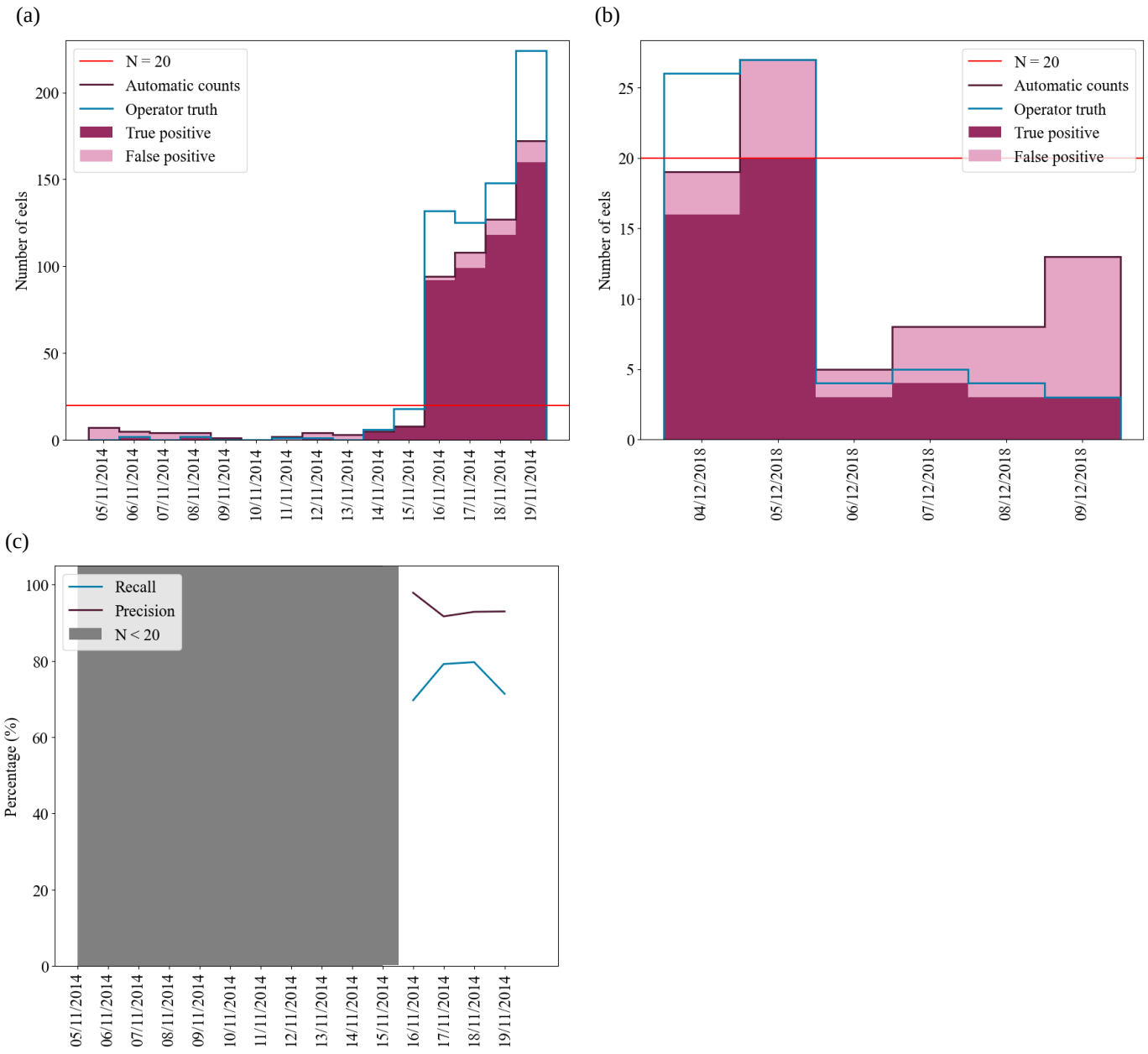


Figure 51 – Number of eels counted by the proposed method compared to the operator truth in the ARIS 1800 videos of Mauzac, with TP and FP displayed, for 15 days in November 2014 (a) and 6 days in December 2018 (b). The number of TP and FP are also displayed as well as the threshold N = 20 corresponding to days where more than 20 eels passages have been automatically counted. Recall and precision achieved for the days of November 2014 are displayed (c).

Regarding the days when the number of daily eels passages is considered low, the recall and precision values are considered not pertinent enough to be calculated. Hence, recall values rely on the operator truth, if only two eels have been counted for one day, missing one will immediately decrease the recall rate to 50%. Similarly, if one eel and one non-eel object is counted by the method, the precision rate is equal to 50%. However, a high amount of FP can be observed

on 09/12/2018 while the number of TP almost fits the operator truth. The visual review of the FP highlights arcing phenomenon in the videos for which 8 out of the 10 FP of 09/12/2018 can be imputed of.

3.2 Mauzac long-term monitoring of European eels

The whole dataset of the three monitored migration periods have been also analysed using the pipeline. No comparison with an operator truth is possible but the aim is to display the counts provided by the automatic method on long-term monitoring periods (Figure 52).

The monitoring is not continuous because of some technical issues that have interrupted the recording during some periods of time, displayed in grey in Figure 52. The European eel's counts are displayed according to the water flow observed at the Lamonzie measurement station, located 40 km upstream the Mauzac monitoring site. For the three monitoring periods, the number of eels simultaneously increases as the water flow increases. It occurs in November 2014, December 2015 and January 2016 when peaks in the number of eels are observed. The same phenomenon is observed in October 2015 and in December 2018 but with lower peaks.

The five peaks mentioned previously occurring from October to January are in concordance with the migration period mainly observed for European eels in France [Righton et al., 2016]. However, eels passages have also been counted in acoustic videos in May 2016 and in June 2016. A visual check of the videos of May 2016 has been made. On the 105 detections counted by the method during this period, only 5 are true European eels. Most of these false detections are due to the continuous presence of arcing phenomenon (Figure 27.a). Their appearance and disappearance during the recordings lead to the appearance of a slender shape which distorts the classification. This amount of false detection was not observed during the periods when a truth operator is available (Figure 51.a and b), as the camera maintenance was focused on the winter period.

However, peaks of European eels counts are not systematically observed when water flow increases in the periods of known migration. In January 2016, a high peak of eels passages is observed but it decreases while the water flow reaches high values. Hence, according to Figure 53, the number of counted eels passages per acoustic video increases as the water flow increases before abruptly dropping above 300 m^3/s. Counts of eels are still provided by the method above 300 m^3/s but it mainly corresponds to the FP of May 2016. The threshold value of 300 m^3/s coincides with the hydropower plant design flow. As previously explained, above this value the telemetry studies show that the European eels are passing on the right bank of the inlet canal rather than on the left one where the AC is installed [ECOGEA, 2014]. Besides, once the hydropower plant design flow is reached, spillage occurs at the dam offering another way of

passage to the European eels. Not seeing eels on camera when these conditions are met does therefore not mean that none of them migrate.

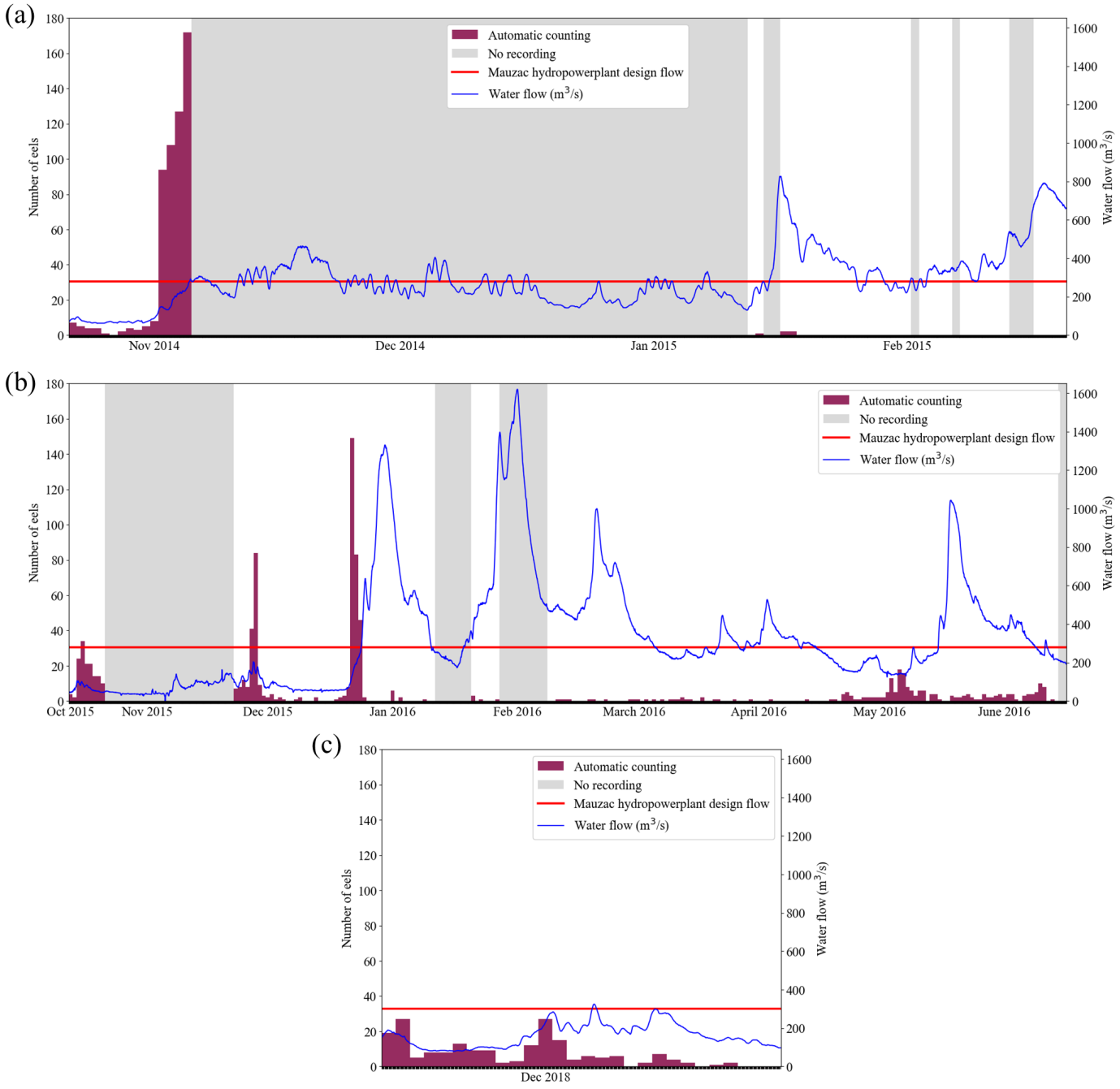


Figure 52 – Number of European eels counted by the proposed method on the ARIS 1800 videos of Mauzac for three migration periods: from November 2014 to early March 2015 (a), from late October 2015 to June 2016 (b), and from December 2018 to January 2019 (c). The water flow of the Dordogne River is also displayed as well as the Mauzac hydropower plant design flow. The periods for which no recording has been done due to technical issues are highlighted in grey.

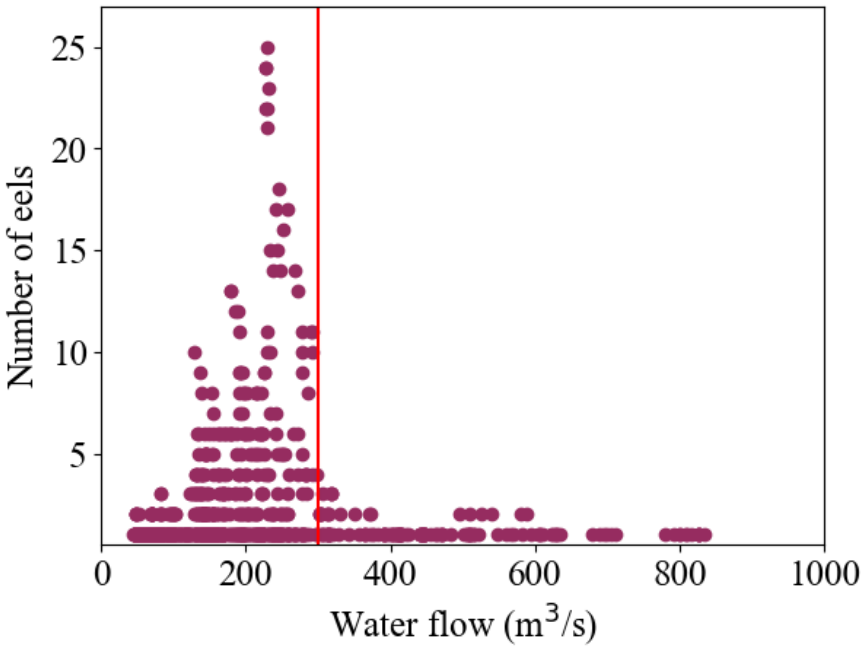


Figure 53 – Number of eels counted by the method in each acoustic video according to the corresponding water flow of the Dordogne River, for the three periods of migration recorded. The red line represents the limit above which spillage starts at the dam.

3.3 Comparison with the model of the European eel's migration prediction of Mauzac

The turbine shutdowns of the Mauzac hydropower plant are planned based on a migration prediction model [E. De Oliveira, personal communication]. As explained in Chapter 1, the model takes into account the water flow of the Dordogne River and of one of its tributaries the Vézère River up to the eight previous days.

Figure 54 displays the prediction of the model for the three migration periods. The peaks of migration observed thanks to the automatic counting of the European eels passages on acoustic videos coincide with the predictions of the model. However, in October 2015 and in December 2015, the eels are also observed on the videos, after and before the predicted period of migration. In December 2018 and in March 2016, eels have also been counted by the methods while no migration has been predicted by the model. Because of the low counts on this period and of the potential errors of the counting method in this case as highlighted in the previous subsection, the difference with the model prediction has to be handled carefully.

The other periods of predicted migration do not fit any peaks of migration observed in the counts of acoustic videos. The water flow of the Dordogne River higher than the hydropower plant design flow of 300 m^3/s explains that no eels have been observed in the acoustic videos.

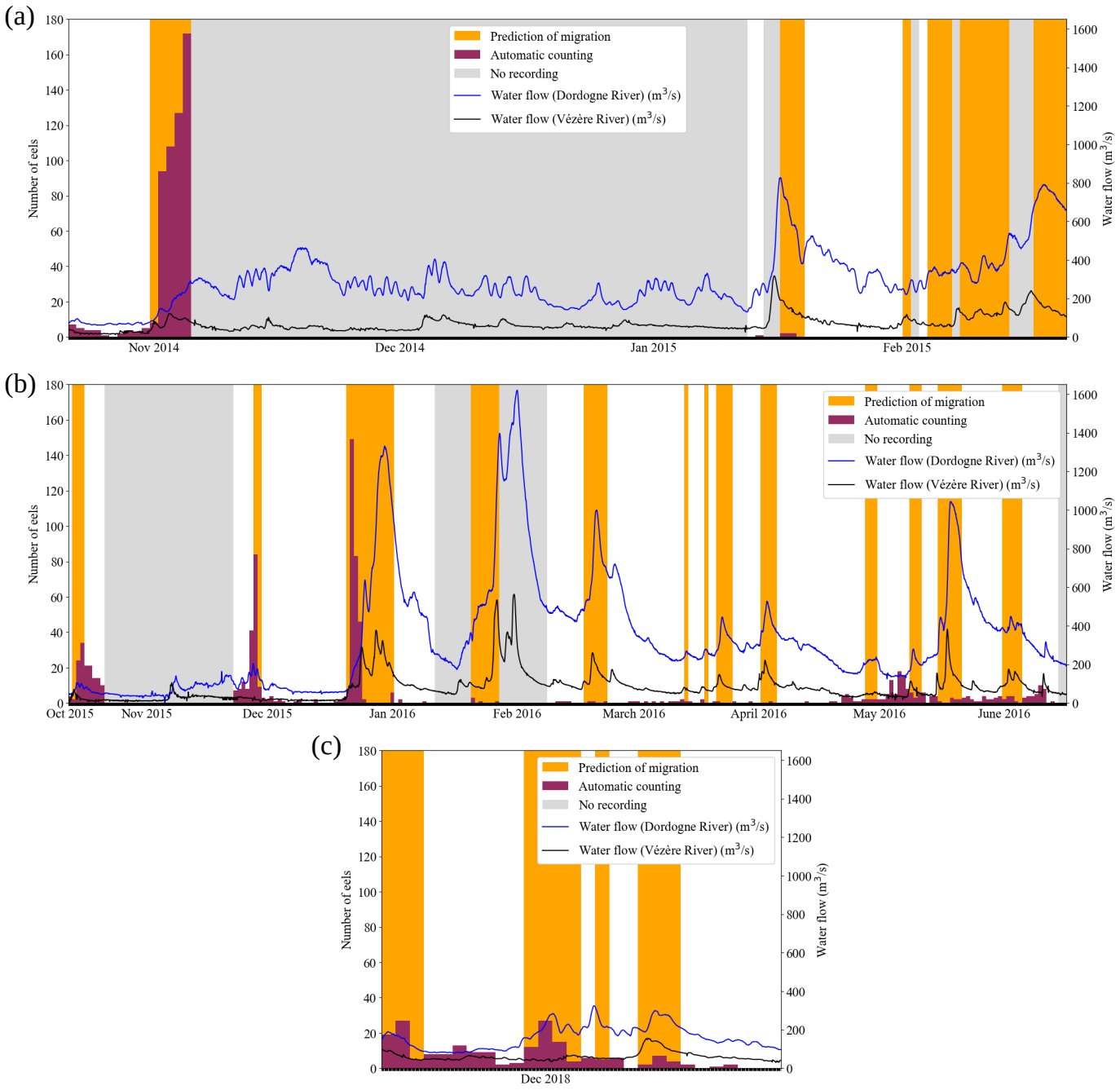


Figure 54 – Number of European eels counted by the proposed method on the ARIS 1800 videos of Mauzac and the corresponding nights when the model predicted a downstream migration (in orange), for three migration periods: from November 2014 to early March 2015 (a), from late October 2015 to June 2016 (b), and from December 2018 to January 2019 (c). The water flows of the Dordogne River (in blue) and of the Vézère River (in black) are also displayed. The periods for which no recording have been done due to technical issues are highlighted in grey.

Besides, from the automatic counts can be deduced the distribution of the passages per day. To reduce the impacts of the FP on this analysis, the counts of May 2016 and June 2016 have been removed as it has been shown that they correspond to FP. Figure 55.a highlights that European eels passages are mainly observed in acoustic videos from 6pm to 6am, corresponding

to 81.4% of the passages in the recorded periods. The small peak of passages at 1pm is due to the FP counted during the periods of May and June 2016 (Figure 55 .b). The distribution of the automatic counts is however coherent with the distribution of the TP (Figure 55.c), highlighting the reliability of these observations that are also in accordance with the modalities of the shutdowns that are made from 6pm to 6am.

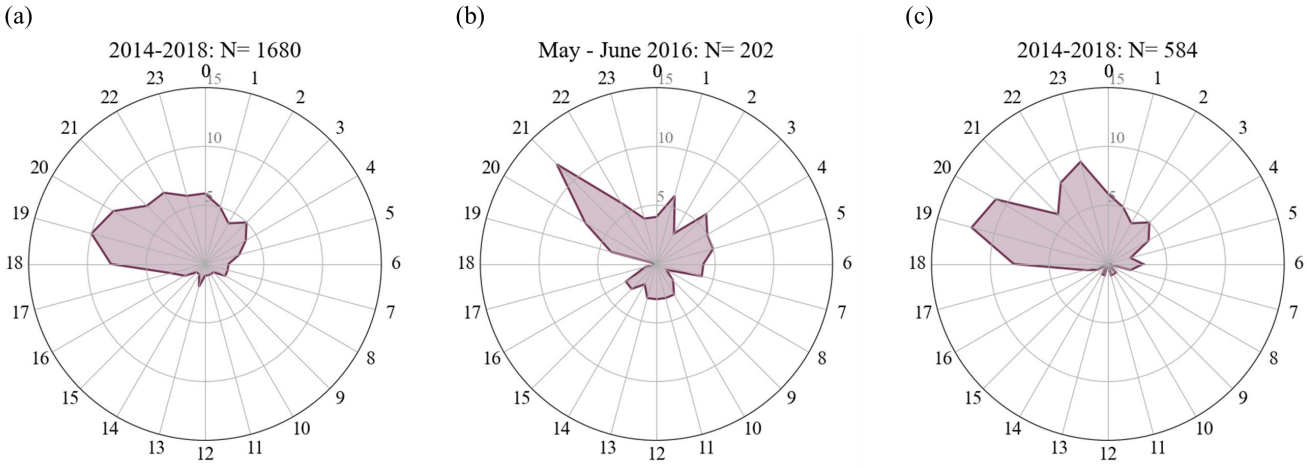


Figure 55 – Hourly distribution of the automatic counts of European eels in acoustic videos from the three recorded periods (N = 1680 passages) (a) and for the period of May to June 2016 (N = 202 passages) (b). Hourly distribution of the TP counts of European eels in acoustic videos for the reviewed videos of November 2014 and December 2018 (N = 584 passages) (c).

3.4 Optimising the monitoring site coverage

To maximise the quantification of eels passages, the use of a BlueView camera providing wider FOV has been investigated. A BlueView camera (BV) has therefore been temporarily installed at Mauzac during four days of the last migration period recorded. Figure 56 displays the number of eels passages automatically counted respectively on ARIS 1800 and BV datasets. According to these results, the BV videos allow us to observe three times more eels passages than in the ARIS 1800 videos. Similar performances are obtained with BV acoustic data and ARIS acoustic data when using the pipeline proposed by [Le Quinio et al., 2023].

However, despite covering almost 4.5 times more water volume than the ARIS 1800, the coverage of BV remains low compared to the width of the inlet canal. It allows us to partially cover (29% of the water section) one strait out of the three composing the inlet canal (Figure 20).

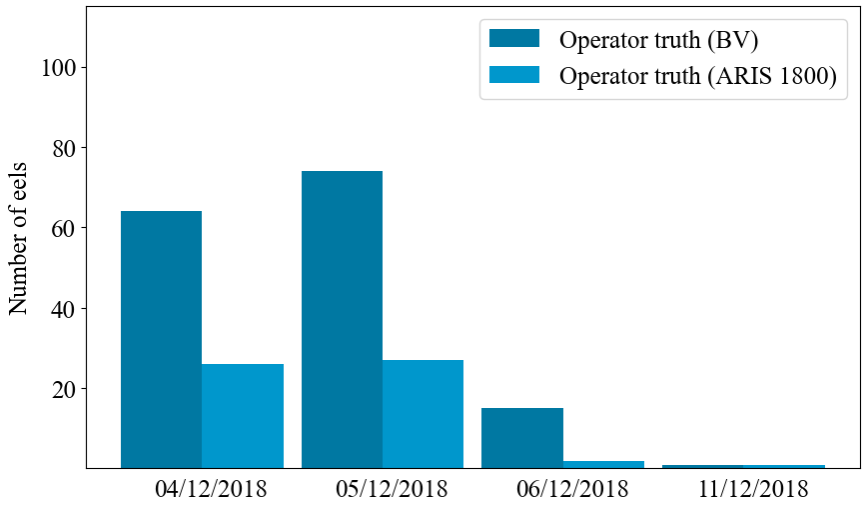


Figure 56 – Comparison of the number of eels counted on the ARIS 1800 and the BV cameras at the Mauzac monitoring site, according to operator truth on four days.

4 EelCounter: an operational and almost real-time tool

Regarding the performance of the proposed pipeline, its implementation as an operational tool was investigated. The aim was to provide a tool that identifies and counts anguilliform fish, on both ARIS and BlueView cameras. To enable its use by a larger number of users without the need of specific informatic knowledge, the pipeline was embedded in a human-machine interface by a contractor, based on the algorithms I developed according to the proposed method. In the frame of this PhD work, program optimization of the pipeline was also done to make my method compatible with a real-time application (Table 9).

Table 9 – Average computation time reached for each of the pipeline steps run in parallel on three processors, using a video of MZC-ARIS and a video of MZC-BV, with a HP ZBook 15 G3, IntelCore i7-6700HQ CPU @ 2.60GHz, 2592MHz, 4 cores, 8 logical processors, 32 Go random-access memory.

Pipeline's step	Average computation time	
	MZC-ARIS n = 4200 frames, t = 10 minutes, 664 x 1276 px	**MZC-BV** n = 2800 frames, t = 10 minutes, 2302 x 1238 px
Step 1: Video conversion to .AVI	~ 4 minutes	~ 4 minutes
Step 2: Echogram calculation (Generic)	~ 3,5 minutes	~ 6 minutes
Step 3: Identification and counting of anguilliform species	~ 2 to 3 minutes	~ 2 to 3 minutes

The echogram calculation step relies on the generic proposed method based on surface area as it performed better while requiring lower computation time. The tracking method is the centroid-to-centroid one but it will have to be replaced by the Kalman filter approach once additional testing is done to strengthen the performance evaluation done in Chapter 2.

The resulting interface allows the operator to get a graphical interface (Figure 57). He can fill in the different elements required for the analysis: the AC model, minimum and maximum ranges of the window length, FOV opening and the value of L_{Target}. Furthermore, the operator indicates the folder containing the raw videos to be analysed and the maximum processors he wants to allocate to the analysis. Progression bar is displayed and updated to inform the operator on which videos are currently processed and at which step of the progress they respectively are. Results of the counting are recorded in a CSV file, one for each day of monitoring, which is progressively updated as videos are analysed. In addition, a graph is displayed in the interface to allow the operator an almost real-time monitoring.

The EDF R&D interface has been tested in laboratory configuration with videos recording in real-time and will be shortly tested in an operational environment.

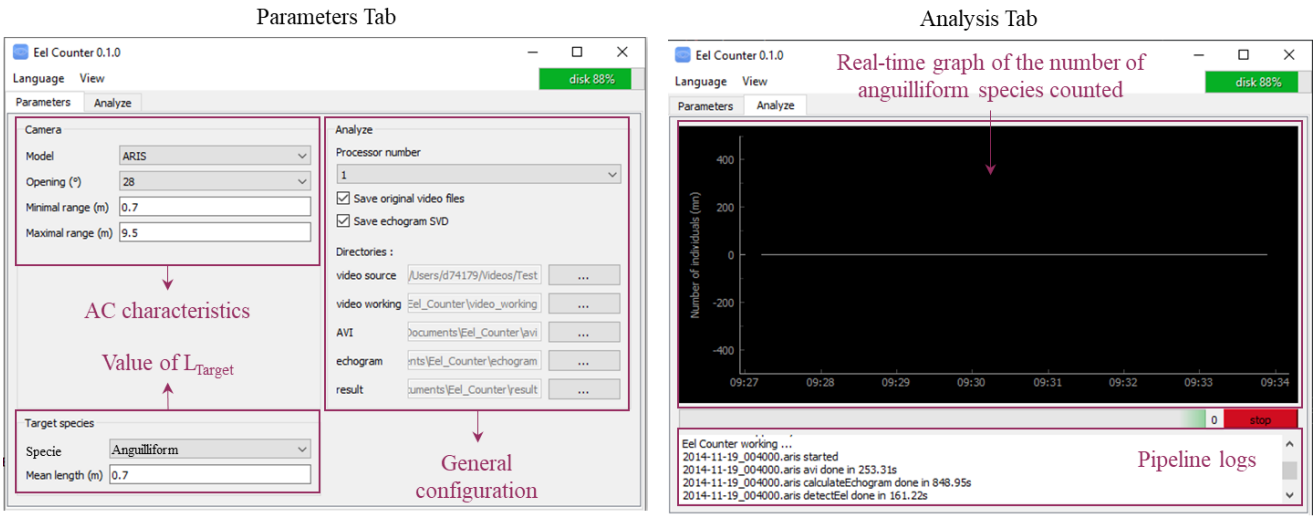


Figure 57 – Snapshot of the *eelCounter* interface. The interface is composed of two tabs: the parameter tab (on the left) that allows the user to fill in its AC characteristics as well as the L_{Target}; and, the analysis tab (on the right) that displays the number of counted individuals.

5 Discussion

5.1 A cross-camera pipeline

The proposed method provides encouraging performances for the identification and count of anguilliform species from AC data. It performs similarly well on MZC-ARIS and MZC-BV

dataset despite the structural noise present on the BV videos. This noise could be explained by the electronic scanning of successive camera beams. This masks the echoes of objects, preventing the method from detecting them as they pass through the noise zone, and thus distorting the tracking algorithm. Processing these frames using the 2D Fourrier transform would make it possible to detect the spatial frequency of this noise and thus correct it. However, because of the large size of the BV frames and the frame rate, this operation would be time-consuming. Nevertheless, these performances highlight the cross-camera ability of the proposed method and the feasibility of having a single pipeline for acoustic videos analysis.

However, the performance reached on the dataset of Port-La-Nouvelle, introduced in [Le Quinio et al., 2023], is lower than expected. The population of European eels of Port-La-Nouvelle is mostly composed of males with smaller size [Lagarde et al., 2020] than the population of Mauzac [ECOGEA, 2014]. These poorer results therefore highlight the relationship between the pipeline performance and the target's size. [Fernandez Garcia et al., 2023] highlight similar conclusions for fish identification using convolutional neural networks: with recall rates decreasing from 65% for eels longer than 60 cm to 22% for those shorter than 60 cm. Hence, targets of small sizes will be defined by fewer pixels compared to larger targets, limiting the morphological information available. The more pixels define the target, the more reliable the morphological analysis is.

In addition to the number of pixels defining a target, the beams and samples resolutions of the acoustic image might also influence the target's image and therefore the pipeline ability to identify it. Cross range resolution relying on target's range, the lower performance at high range coincides with a poorer beam resolution. The lower performance at small range can be explained by recurrent errors due to arcing phenomenon already observed in the literature [Mueller et al., 2008, Burwen et al., 2010]. These phenomena are never observed on the BlueView cameras. They are caused by fish or highly reflective targets moving near the camera and might be more frequent when the ARIS lens box is lined with silt, and consequently can be an indication of the need of a maintenance of the ARIS cameras.

Another factor that may influence the pipeline performance, not yet investigated, is the orientation of the target towards the acoustic beams of AC as it highly affects the reconstructed target's image and therefore its morphological characteristics. This point will be discussed in the general discussion of the manuscript.

Despite all of these elements to be taken into account and which are influencing its performance, the proposed pipeline provides encouraging identification and counting of large anguilliform species, such as the European eel observed in the Dordogne River at Mauzac. Its operational use has therefore been carried out with some optimisations of the initial pipeline

in order to reach an almost real-time calculation. In addition to saving computation time, the echogram calculation based on surface area does not affect the performance, reaching better results as explained in Chapter 2. Another way to strengthen the whole pipeline could be to remove the empirical thresholds and carry out the descriptors classification through machine learning classifiers.

5.2 AC as an operational tool thanks to the automatic and real-time analysis

The AC combined with an almost real-time method to automatically analyse their videos provide continuous monitoring of eels passages during long periods of time. However, the use of counting provided by the automatic analysis of acoustic videos must take into account the specificities of the monitoring site and the location of the AC. Hence, at the Mauzac monitoring site, AC is covering a limited portion of the inlet canal width (10 meters out of the 50 meters) giving only a proxy of the eels passages count. If the location of AC has been defined thanks to radio telemetry and fisheries campaigns to optimise eels passages visualisation, the covered section is not optimal in all environmental configurations. Hence, at high water flow, the radio telemetry highlights that eels are preferentially passing on the opposite bank of the AC [ECOGEA, 2014]. Besides, at high water flow, spillage can be engaged at the dam, opening a new preferential way for eels. High water flow might also change the swimming behaviour of the eels, that might swim deeper in the water column and therefore no longer pass through the AC FOV. High water flow also increase eel's velocity that might be captured on less frames making them more difficult to track and identify.

Location of the AC is therefore of key importance to bring exploitable and reliable information on European eel's passages.

While keeping in mind these limitations, peaks of migration can be observed through acoustic videos providing valuable indication for facility management such as the beginning of the European eel's migration. Although the automatic analysis of the BlueView videos takes a little bit longer than their duration, the delay is short and still allows an operator to get the information needed for the management of the hydropower plant. It also allows the operator to compare the predictions of the downstream migration model to *in situ* observation of passage, in case of low water flow conditions. However, the limited covering of all ways of passages of eels may exacerbate the difficulty to estimate downstream European eels migration fluxes from acoustic videos.

Maximising AC coverage is therefore the way to overcome these limitations. The newest types of AC, such as the BlueView allow a larger water volume coverage, while not requiring maintenance as it would be for the ARIS models. However, the most efficient way to cover a larger

section of the study area is to equip the monitoring site with several AC. As part of the monitoring of the European eel's downstream migration at the Mauzac monitoring site, a second BlueView camera has been installed in 2022 to cover the first section of the inlet canal (Figure 58), 20 metres from the first AC. The installation of two BlueView cameras, on the first and third sections, is highly facilitated by the automation of the analysis, the review by an operator of the two distinct sets of data from two cameras being too time-consuming.

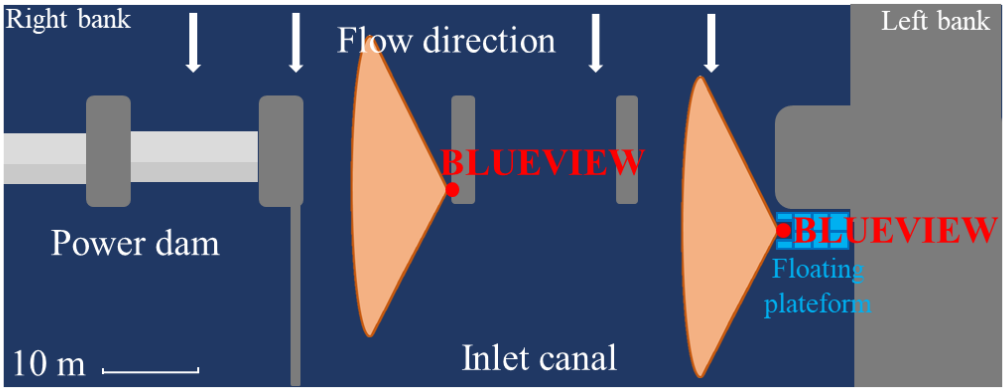


Figure 58 – Scheme of the installation of the two BlueView cameras at the Mauzac monitoring site from 2022 for the monitoring of the European eel's downstream migration.

Through the design of an automatic cross-camera method of counting and identification of anguilliform species embedded in a user-friendly application, we aim to provide a pipeline for anguilliform species identification and an operational tool to complete it, that is not intended to be open-source.

5.3 Morphological characteristics for multi-species classification

Although the morphological analysis method was initially developed for anguilliform species, the feature extraction steps are also applicable to other species, and in particular the shape characterisation metric proposed in this manuscript. However, the information on morphology and movement may not be sufficient for multi-species discrimination. Many species do not have a particularly characteristic shape.

Chapter 4. Swimming behaviour: a complementary tool for species identification

Contents

Il y a une limite à toute chose,
et il faut toujours la dépasser.
Georges Guynemmer

1 Introduction

The deformation of the body of a fish along its movement is the second characteristic of interest that can be extracted from the AC data in order to discriminate species. In addition to the morphological and motion characteristics analysed thanks to the method developed in Chapter 3, this could help us to improve species identification ability. To extract corresponding information from the videos, the deformation of the fish body must be tracked and quantified at each time step. It means that for a same point of the body we want to measure its position gap between two successive time steps, independently from the displacement of the fish. To do so, we use deformation models.

Deformation models are curves or surfaces defined from an input image. They can move under the influence of internal forces coming within the model itself and external forces computed from the processed image [Xu et al., 2009]. In other words, it means that deformation models are designed to be drawn by external constraints from the processed image that command the deformation, while preserving the internal initial shape set up from the input image. Without going into details at this point, we can take an example from [Schaefer et al., 2006] which illustrates this concept with three examples of deformation of an input image (Figure 59). The objective is to deform the image so that the controlling points move from their initial positions (Figure 59.a, in blue) towards their final positions (Figure 59.b-d, in blue), which correspond to external constraints applied to the image. Depending on the type of transformations used, the internal forces that command the deformation differ and so produce the deformed images (Figure 59.b-d).

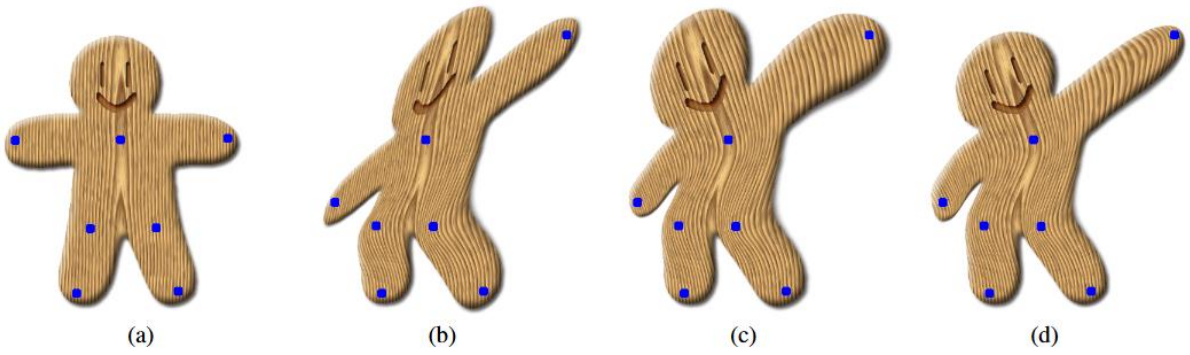


Figure 59 – Examples of moving Least Squares deformations of an initial image (a) based on controlling points (in blue) using affine transformations (b), similarity transformations (c) and rigid transformations (d), from [Schaefer et al., 2006].

Deformation models are a wide area of study. From image segmentation [Xu et al., 2000] to object tracking [Zhong et al., 2000, Dambreville et al., 2006], edge detection and shape modelling [Xu et al., 2009], deformation models are used in plenty of image processing applications. First introduced by [Terzopoulos, 1987], various deformation models have since been proposed in the

literature [Jain et al., 1998, Xu et al., 2000, Xu et al., 2009, Selim and Koomullil, 2016]. However, only some of them might be useful for our study. Two categories of deformation models exist: on one side, geometric models; on the other side, parametric models. Geometric models track the deformation using geometric measures, independently from the parameterisation, allowing the curves and surfaces to be implicitly represented [Mesejo et al., 2014]. Despite their simple implementation and efficient computation time [Gibson and Mirtich, 1997], geometric models do not rely on physical principles to carry out the process, preventing natural deformations. On the contrary, parametric models are explicitly representing curves and surfaces. It uses prior information on their shape [Jain et al., 1998], ensuring a direct interaction with the model [Mesejo et al., 2014]. Among the parametric methods, several physically-based models are able to model complex deformation processes based on the physical rules that govern the object [Gibson and Mirtich, 1997], such as elasticity theory. The most common one is the Mass-Spring Model (MSM) which represents the object as a collection of points of a certain mass that are connected with each other through springs and subjected to elastic forces [Wang et al., 2014]. Drawbacks of this model are its slow convergence and the need for a careful tuning of its parameters [Weng et al., 2006]. Finite-Elements Methods (FEM) have the advantage of providing a more realistic deformation than MSM [Gibson and Mirtich, 1997]. FEM model the object into elements joined at discrete points that are called nodes. A function is calculated for each element to solve the continuous equilibrium equation [Sathyanarayanan et al., 2013] to carry out the deformation. Despite giving more realistic deformation, FEM also present some limitations as they require high computation time restraining their applicability to interactive deformation applications [Weng et al., 2006].

A more recent method is the gradient domain mesh deformation that aims to overcome issues previously observed with physically-based models while presenting many advantages [Xu et al., 2009]. This method is a surface-based, variational optimisation method, i.e., a method that carries out the optimisation over a space of functions. The method explicitly considers the preservation of the local shape properties during the deformation process, overcoming the issue of their distortion generally encountered in other mesh-based models [Xu et al., 2009]. To do so, it directly encodes the geometric details of the object by the Laplacian calculated on the curve's discrete points [Xu et al., 2009]. During the deformation process, the preservation of the Laplacian coordinates thus ensures the preservation of the mesh details [Xu et al., 2009]. This deformation process is seen as an energy minimisation, as for many of the parametric models [Xu et al., 2000], including internal and external terms. Internal energy terms aim to preserve the local shape properties, so that the deformation of the mesh is carried out to minimise the divergence from the initial one. External energy terms oppositely ensure that the mesh deforms to the desired output positions. To solve this energy minimisation problem, nonlinear optimisation methods are usually preferred as it allows us to compose with nonlinear constraints to induce expected deformations [Xu et al., 2009].

Those gradient domain techniques have been successfully used in different studies for mesh deformation [Xu et al., 2007, Weng et al., 2006, Huang et al., 2006]. An example of their applications by [Weng et al., 2006] is displayed in Figure 60.

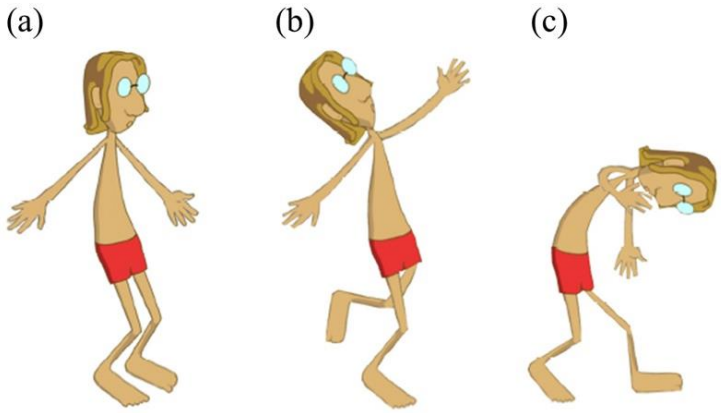


Figure 60 – Example of two deformations (b) and (c) of a 2D cartoon character image (a), from [Weng et al., 2006] works. The deformations have been carried out using gradient domain techniques and nonlinear least squares optimisation.

Deformation model implementation is the first step to describe the deformation of the fish body. The second step to be carried out is the processing of this deformation to extract characteristics useful for species discrimination.

In a first approach, we can see the problem as a succession of 1-D waves being propagated. Works have been conducted on the modelling of deformation waves [Engelbrecht and Berezovski, 2015]. However, it is a complex approach that may not be adequate for real-time analysis. Dynamic Time Warping (DTW) could be another way to deal with our 1D-data (Figure 61). If we consider our deformation data as time-series, then DTW can measure similarity between them, pairs by pairs, by allowing elastic transformation to minimise shifting and distortion effects [Senin, 2009].

In a second approach, we can see our deformation data as a 2-D matrix, that we call "map of deformation", and apply image processing methods to extract discriminating information. A lot of image processing tools are available for this kind of analysis. However, some of them might be more adequate. Our body deformation data can be considered periodic with species specific patterns. Fast Fourier Transform as well as auto-correlation are commonly used in the literature to study periodic structural patterns, respectively in the frequency and coordinate spaces [Li, 2019] and might thus help to highlight patterns present in AC data. Auto-correlation is a measurement of similarity between two signals as a function of the displacement of one relative to the other. Image matching descriptors can also be considered as they extract features of interest from images [Li, 2017, Karami et al., 2017]. The Scale-Invariant Feature Transform

(SIFT) image matching descriptor introduced by [Lowe, 2004] aims to detect and describe key points of an image that are invariant to various images transformations (Figure 62.a). The Speeded Up Robust Features (SURF) method is inspired by the SIFT method and has been developed by [Bay et al., 2006] in order to be quicker than its predecessor (Figure 62.b). The main difference occurs at the key points detection step for which SURF uses another way than SIFT to approximate one of the mathematical operators needed [Bay et al., 2006]. Finally, the Oriented FAST and Rotated BRIEF (ORB) method developed by [Rublee et al., 2011] has the same structure as the two other methods (Figure 62.c). It first detects key points using the FAST corner detection algorithm [Calonder et al., 2011] before quantitatively describing them using the BRIEF algorithm [Viswanathan, 2011]. Several studies have been conducted to compare the three methods. They highlight the faster computation time of ORB which performs as well as SIFT in many situations [Rublee et al., 2011], both methods outperforming the SURF one.

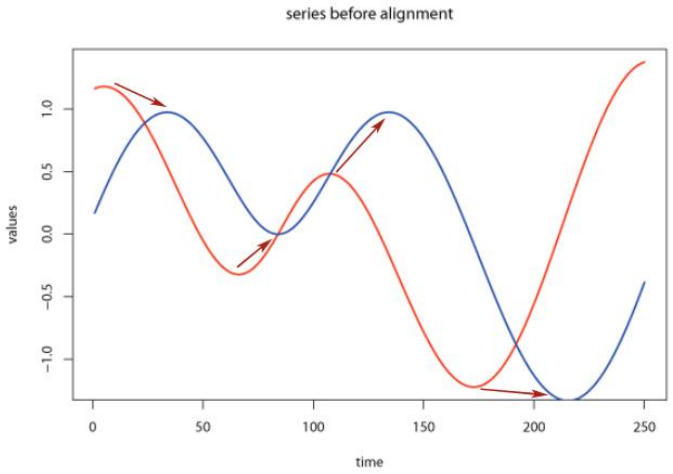


Figure 61 – Example of two time-series before aligning, with the expected points of match (in red arrow) using DTW, from [Senin, 2009]

All of these methods are descriptors that are used in combination with classifiers to carry out the classification process. In this chapter and similarly as in Chapter 3 for morphological and motion characteristics, we do not seek to develop a model of classification on fish body deformation. The aim is to later develop a global classification model with all of the information that has been extracted from AC data. However, at this step of our development, we need to get an indication on the discrimination ability of the information we extract. Information is retrieved using the imaging matching processors on the map of deformation calculated from the fish body. We therefore aim to compare the maps of deformation between each other and give them scores of similarity. Matching methods such as the one proposed by [Lowe, 2004] through the SIFT descriptor, that performs well on images with varying intensity values as ours [Karami et al., 2017], might help us to provide features matching and give us enough indication

to produce a score of similarity.

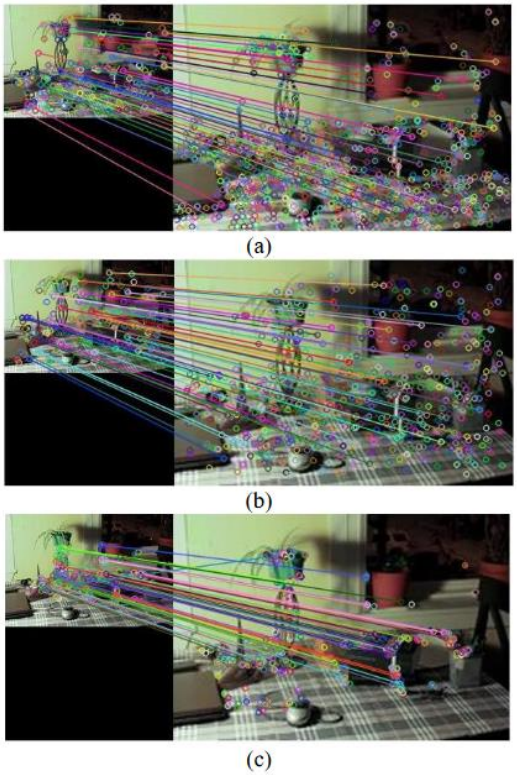


Figure 62 – Examples of the matching of two different scaled images, using SIFT (a), SURF (b) and ORB (c) methods, from [Karami et al., 2017]. The circles are the key points extracted by the methods. The matching of the points from one image to another is made according to the features of these key points extracted by each of the three methods.

In this chapter, we thus propose a method that combines the calculation of the fish body deformation using a deformation model with image processing tools to provide similarity scores to discriminate species.

2 Method

2.1 Characterising the target deformation over time

2.1.1 Generation of a deformable grid

To use a deformation model, we first need to generate the grid that must be deformed. It is generated as a grid of triangular mesh using the contour's points of the binary shape given as input through the function *triangle* of the *MeshPy* package.

All the coordinates' points of the grid are summarised in the scalar V. From the grid, we differentiate the N points on the outer side (in blue in Figure 63), defining V_b, from the M

points of the inner part of the grid (in purple in Figure 63) defining V_{inner}. We therefore have the N first points of V corresponding to V_b and the M following points corresponding to V_{inner}. The vector V_b is sorted in such way that the neighbours of $v_i \in V_b$ in the outer curve of the grid are v_{i-1} and v_{i+1} (Figure 63). Regarding V_{inner}, a 2-column matrix e is defined corresponding to all the combinations of edges $v_n - v_m \in V_{inner}$. Thus, if v_n and v_m form the z^{th} edge of the inner part of the grid, then $e_{ED_z} = (v_n, v_m)$.

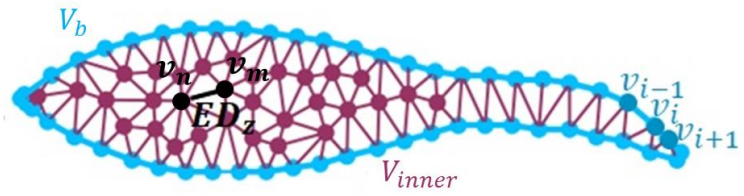


Figure 63 – Scheme explaining the definition of V_b and V_{inner} scalars, and of the column-scalar *ED*. The blue points correspond to the outer side and compose the scalar V_b. The pink points correspond to the inner part and compose the scalar V_{inner}. The black line corresponds to the z^{th} edge of the inner part of the grid that composes the column-scalar *ED*.

2.1.2 Deformation model

Deformation model aims to modify the mesh according to the constraints set up. These constraints are wrapped in the energy of deformation. By minimising it, the deformation process is carried out. We based our model on the energy of deformation (eq. 45) proposed by [Weng et al., 2006]. The three first terms of (eq. 45) aim to preserve the local properties of the mesh while the last one aims to give the direction of the deformation.

$$\|LV - \delta(V)\|^2 + \|MV\|^2 + \|HV - e(V)\|^2 + \|CV - U\|^2 \qquad \text{(eq. 45)}$$

With, $\|LV - \delta(V)\|^2$ aiming to preserve the global aspect of the grid's curve; $\|MV\|^2$, aiming to preserve the local areas inside the grid through the mean values coordinates; $\|HV - e(V)\|^2$, aiming to preserve the local areas inside the grid through edge lengths; and, $\|CV - U\|^2$, represents the constraints on the positions for which we want our model to deform. All these terms are explained below.

The first term $\|LV - \sigma(V)\|^2$

Minimising $\|LV - \sigma(V)\|^2$ helps to preserve the global aspect of the grid's curve. From [Weng et al., 2006], we have :

$$\delta_i = \mathcal{L}_p(v_i) = v_i - (v_{i-1} + v_{i+1})/2 \qquad \text{(eq. 46)}$$

With $v_i \in V_b$; and, v_{i-1} and v_{i+1}, the two neighbours of v_i in the curve.

So, we can deduce:

$$\forall i \in [1, N], \forall j \in [1, N+M], L_{i,j} = \begin{cases} 1 & \text{if } i = j \\ -0.5 & \text{if } i = j-1 \text{ or } i = j+1 \\ 0 & \text{else} \end{cases} \quad \text{(eq. 47)}$$

The second term $\|MV\|^2$

Both second and third terms of the energy of deformation aim to preserve the local areas inside the grid. Minimising $\|MV\|^2$ allows us to preserve the position of $v \in V_{inner}$ according to its neighbours. From [Weng et al., 2006],we can deduce (eq. 48) and (eq. 49):

$$\forall i \in [N+1, N+M], \forall j \in [1, N+M], M_{i,j} = \begin{cases} M_{i,j} = -w_{i,j} & \text{if } j \in T(i) \\ 0 & \text{else} \end{cases} \quad \text{(eq. 48)}$$

$$\forall i \in [N+1, N+M], M_{i,i} = -\sum_{j \in T(i)} w_{i,j} \quad \text{(eq. 49)}$$

With *T(i)*, all of the points in the neighbourhood of v_i, meaning the points being part of the triangles for which v_i is one of the summit (Figure 64); and $w_{i,j}$, the mean value coordinates of v_i with v_j in the area formed by its neighbours (Figure 64) and calculated from [Floater, 2003]:

$$w_{i,j} = \frac{tan(\alpha_j/2) + tan(\alpha_{j+1}/2)}{|v_i - v_j|} \quad \text{(eq. 50)}$$

With α_j, the angle $\widehat{v_j v_i v_{j+1}}$; α_{j+1}, the angle $\widehat{v_{j+1} v_i v_{j+2}}$; and v_{j+2}, the next point in *T(i)* following v_j (Figure 64).

The third term $\|HV - e(V)\|^2$

Concerning the third term $\|HV - e(V)\|^2$, its minimisation allows us to maintain the edge length between two points v_n and v_m from V_{inner}. From [Weng et al., 2006], we have:

$$\sum_{(i,j) \in E_g} \|(v_i - v_j) - e(v_i, v_j)\|^2 \quad \text{(eq. 51)}$$

With:

$$\forall s \in [1, S], e_s = e(v_{ED_{i,1}}, v_{ED_{i,2}}) = \frac{\widetilde{l}_{ED_{i,1}, ED_{i,2}}}{l_{ED_{i,1}, ED_{i,2}}} (v_{ED_{i,1}} - v_{ED_{i,2}}) \quad \text{(eq. 52)}$$

With $\widetilde{l}_{ED_{i,1}, ED_{i,2}}$, the distance, before any deformation, between the points of the i^{th} edge whose indices in V_{inner} are respectively $ED_{i,1}$ and $ED_{i,2}$; and, $l_{ED_{i,1}, ED_{i,2}}$, the same distance

but calculated at the current deformation step. From (eq. 51), we deduce matrix H:

$$\forall i \in [1, S], \forall j \in [1, N+M], H_{i,j} = \begin{cases} 1 & \text{if } ED_{i,1} = j \\ -1 & \text{if } ED_{i,2} = j \\ 0 & \text{else} \end{cases} \quad \text{(eq. 53)}$$

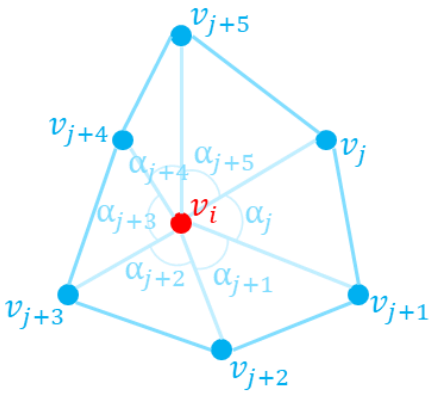


Figure 64 – Neighbourhood of the point v_i in red, all the points in blue are part of *T(i)*. Angles α are also represented in light blue.

The fourth term $\|CV - U\|^2$

The fourth and last term $\|CV - U\|^2$ represents the constraints on the positions for which we want our model to deform. The matrix corresponds to the coordinate of points we want the corresponding grid's points to tend to. In our method, we only give instruction on the outer points V_b of the grid (eq. 54).

$$\forall v_i \in V_b, v_i = u_i \quad \text{(eq. 54)}$$

With u_i, the constraint given at the i^{th} points of the grid's curve.
From (eq. 54) we can deduce both (eq. 55) and (eq. 56):

$$\forall i, j \in [1, N+M], C_{i,j} = \begin{cases} 1 & \text{if } i < N \text{ and } i = j \\ 0 & \text{else} \end{cases} \quad \text{(eq. 55)}$$

$$\forall i \in [1, N+M], U_i = \begin{cases} u_i & \text{if } i < N \\ 0 & \text{else} \end{cases} \quad \text{(eq. 56)}$$

All of the matrices of the deformation model are initialised before the first iteration of the deformation process. Matrices L, M, H and C do not require to be updated later in the process while matrices σ, ED, e and U are re-defined at each iteration according to the value of V at the new iteration process.

2.1.3 Setting up the position constraints

The definition of the matrix U of the deformation between t=k-1 and t=k relies on the contours' coordinates $\mathcal{C}_k$ at t=k of the image to be deformed to. From all the contour's points $\mathcal{C}_k$, we must find the ones that correspond to the curve's points V_b of the mesh. On the one hand, the curvilinear distance of each point of V_b to the mesh's origin, and on the other hand, the curvilinear distance of each point of $\mathcal{C}_k$ to the contour's origin, are calculated according to (eq. 57) (Figure 65) respectively.

$$l_{A,i} = \sum_{j=1}^{i} \|p_{j-1} - p_j\|_2 \qquad \text{(eq. 57)}$$

With $l_{A,i}$, the curvilinear distance of the i^{th} points of the curve A; and, p_{j-1} and p_j, two neighbouring points of the curve A.

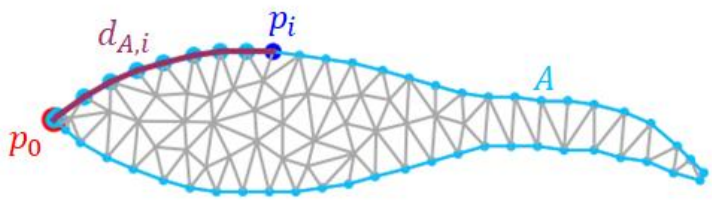


Figure 65 – Graph showing the curvilinear distance $d_{A,i}$ in pink for the i^{th} point p_i in dark blue of the curve A in light blue.

From the values of the curvilinear distances we can identify the matching points on $\mathcal{C}_k$ (Figure 66) and thus deduce the matrix U (eq. 58) of the deformation from t=k-1 to t=k.

$$\forall\ v_i \in V_b,\ U_i = \mathcal{C}_j \text{ if } l_{\mathcal{C},j} = l_{V_b,i} \qquad \text{(eq. 58)}$$

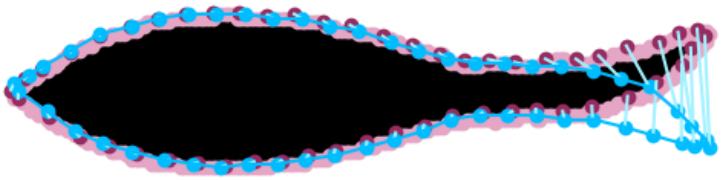


Figure 66 – Matching of the curve's grid at t=k-1, in blue, with the corresponding points in the contour of the image at t=k+1, in pink.

2.1.4 Initialisation of the deformation process

Each process of deformation is carried out once the whole tracking of the targets is completed. The grid is initialised based on the target's successive binary images, prior to the first deformation iteration. The shape used for the initialisation influences the grid definition. To get a generic method and decorrelate the grid's definition from the target to deform to, we first generate the grid from an ellipsoidal shape (Figure 67.a). The ellipsoidal shape is equal to the maximum of the main and minor axis length of the target's images. After being defined (Figure 67.a), the grid is deformed to fit the image of reference of the target (Figure 67.b and Figure 67.c). The image of reference is chosen to be the most representative of all of the images of the target. To do so we empirically decided to choose as the most representative image the one whose length is equal to the 3^{rd} quartile of all of the images length. The resulting coordinates of the mesh define V_0 which can then be used to initialise the deformation process, meaning to initialise all of the matrices previously introduced. The deformation from the detection at t=k is then carried out from the grid position at t=k-1, and so on until we reach the last detected image of the target.

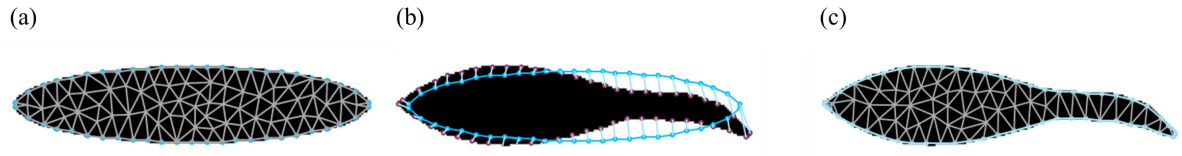


Figure 67 – The successive steps for the model initialisation; (a) the grid generation from the ellipse shape, (b) the deformation of the grid to fit the contours of the image of reference, (c) the deformed grid used to initialise the matrices of the energy of deformation.

It is important to notice that a deformation model is not designed to handle big differences such as translation or rotation from one iteration to another. Such big changes, if existing, must be handled in a separate way prior to the deformation process. First, the global orientation of the object with the horizontal axis is calculated using the dedicated attribute of the *regionprops* function of the *scikit-image* library. This orientation θ corresponds to the ellipse that fits the object the best. A rotation of the object is carried out (Figure 68.a and Figure 68.b) using the corresponding matrix R (eq. 59). A translation is then completed to match the centroid of the object c_t with the one of the previous time step, c_{t-1} (Figure 68.b and Figure 68.c).

$$R = \begin{pmatrix} cos\ \theta & -sin\theta & (1 - cos\ \theta) \cdot c_x - sin\ \theta \cdot c_y \\ -sin\ \theta & cos\ \theta & sin\ \theta \cdot c_x + (1 - cos\ \theta) \cdot c_y \end{pmatrix} \qquad \text{(eq. 59)}$$

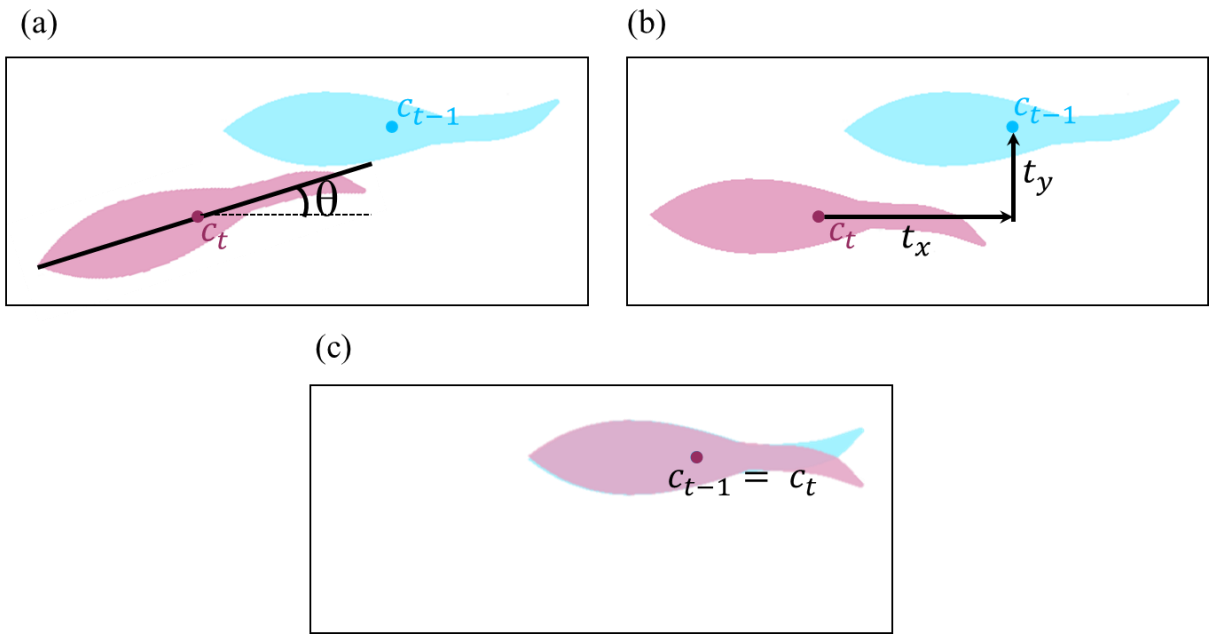


Figure 68 – (a) Calculation of the θ angle, (b) Rotation of the object at t=k in pink to fit the horizontal axis, and calculation of the values t_x and t_y, (c) translation of the object at t=k according to t_x and t_y.

2.2 Discrimination from the maps of deformation

2.2.1 Quantification of the deformation

Prior to quantifying the deformation, a maximisation of the common area of the object at t=k and t=k-1 is carried out by translation. The aim is to avoid wrongly quantifying deformation that would have been due to a lack of precision during the handling of the displacement at the previous step of the process. Deformation can then be quantified between V_{k-1} and V_k by calculating the scalar d (eq. 60) (Figure 69). It represents the distance of the curve points at t=k-1 with the corresponding ones at t=k.

$$\forall\, i \ \in [1, N], d_i = v_i^k - v_i^{k-1} \qquad \text{(eq. 60)}$$

With v_i^k and v_i^{k-1}, the coordinates of the points, respectively at t=k and t=k-1, from which the deformation is calculated.

Calculating d at each iteration generates a matrix D that we call our map of deformation (Figure 70). The vertical axis corresponds to the number of detections, i.e., to the time axis, while the horizontal axis corresponds to the upper curve points from the head to the back and to the lower curve points from the back to the head of the target.

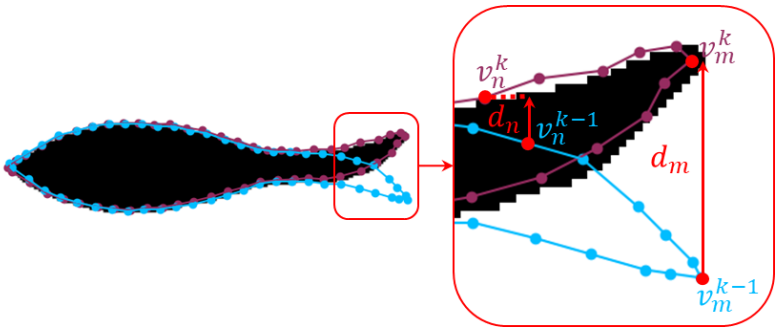


Figure 69 – Examples of two distance d_n and d_m between two sets of points.

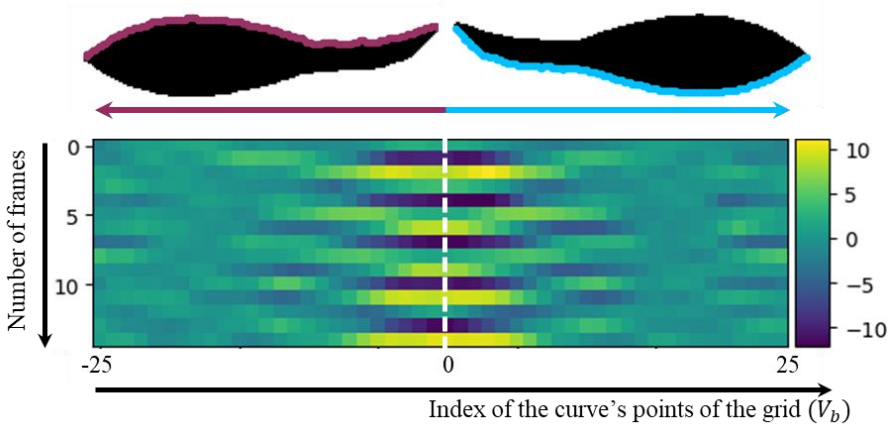


Figure 70 – Example of a map of deformation displaying the deformation values calculated for each point of the curve's grid, at each detection of the target. The vertical axis corresponds to the time axis. The horizontal axis corresponds to the number of outer points of the deformation model. The first half on the left side corresponds to the points of the upper side of the fish silhouette, from the head to the tail as represented in purple on the left fish. The second half corresponds to the points of the lower side of the fish silhouette, from the tail to the head, as represented in blue on the right fish. The colour bar on the right of the map of deformation corresponds to the range of the deformation values, in pixel.

2.2.2 Signal processing of the maps of deformation

Because of the frequency of acquisition of the acoustic cameras, the period of time between two successive detections is lower than for optical cameras (7 to 8 frames per second in average for our AC data against 24 frames per second for optical cameras) and might thus prevent the detection of some deformation patterns. A linear interpolation (eq. 62) on m points (eq. 61) along the time axis, *i.e.* along the vertical axis of the map of deformation (Figure 73.a), is done to reach a frame rate close to the one of optical cameras.

$$m = N \cdot \frac{24}{FPS} \qquad \text{(eq. 61)}$$

With N, the initial number of frames in the maps of deformation and FPS, the frame rate of the acoustic video.

y, the interpolation of the coordinates along the vertical axis is then defined by:

$$y = y_{i-1}(\frac{x_i - x}{x_i - x_{i-1}}) \ + \ y_i(\frac{x - x_{i-1}}{x_i - x_{i-1}}) \qquad \text{(eq. 62)}$$

With x, the coordinates along the horizontal axis of the points to interpolate; x_{i-1}, y_{i-1}, x_i, y_i, the coordinates of the two known points between which the point to interpolate is.

To filter noise and highlight major patterns of the map of deformation, we calculate the correlation of the deformation signal at each time step with a Gaussian (Figure 71) whose function is:

$$f(x) = e^{-\frac{(x-\frac{N}{2})^2}{2\sigma^2}} \qquad \text{(eq. 63)}$$

With N, the size of the map of deformation along the horizontal axis, and σ, the standard deviation of the Gaussian.
We choose to use a small value for σ, $\sigma = \frac{3}{2}$.

Last, the map of deformation is decomposed using Singular Value Decomposition before being reconstructed using the two first eigenvectors. It allows us to discard the secondary eigenvectors and focus on the predominant ones that explain more than 90% of the information, on average, of the maps of deformation of the whole dataset (Figure 72).

The combination of the three signal processing steps allows us to get a cleaner signal to process (Figure 73).

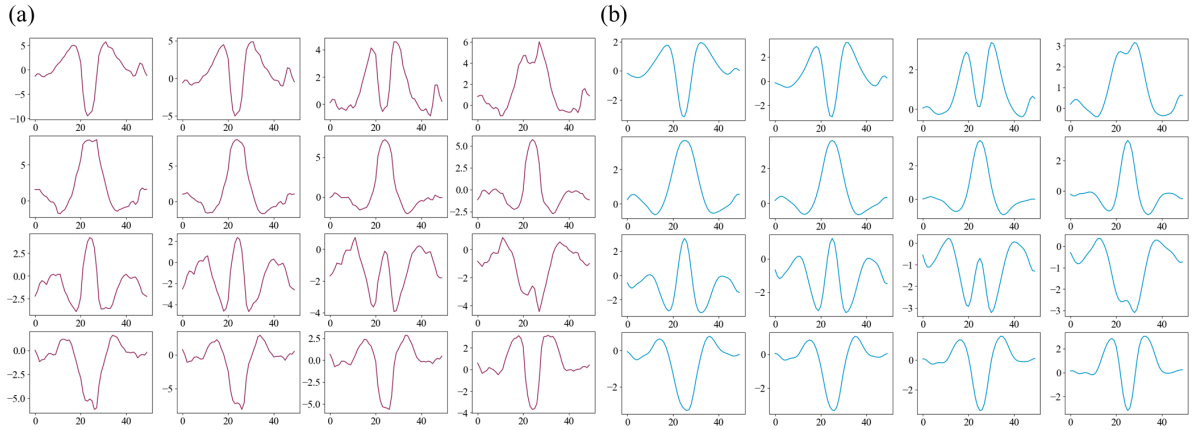


Figure 71 – Examples of successive deformation signals at each time step of a European catfish body deformation process. The original signal after linear correlation is displayed in (a), its post-processing after correlation with a Gaussian is displayed in (b). The horizontal axis corresponds to the 50 outer points of the deformation grid. The vertical axis corresponds to the value of the deformation calculated as a distance in pixels.

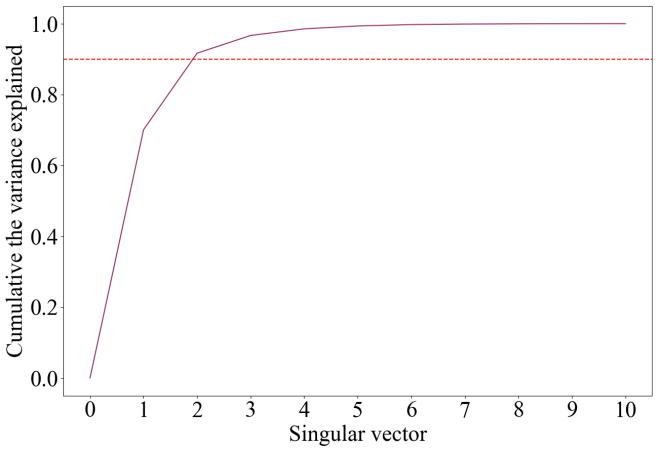


Figure 72 – Averaged cumulative explained variance for the 10 first eigenvectors of the SVD carried out on each map of deformation of the 31 individuals of the dataset. The dotted red line corresponds to the threshold of 90% of explained variance.

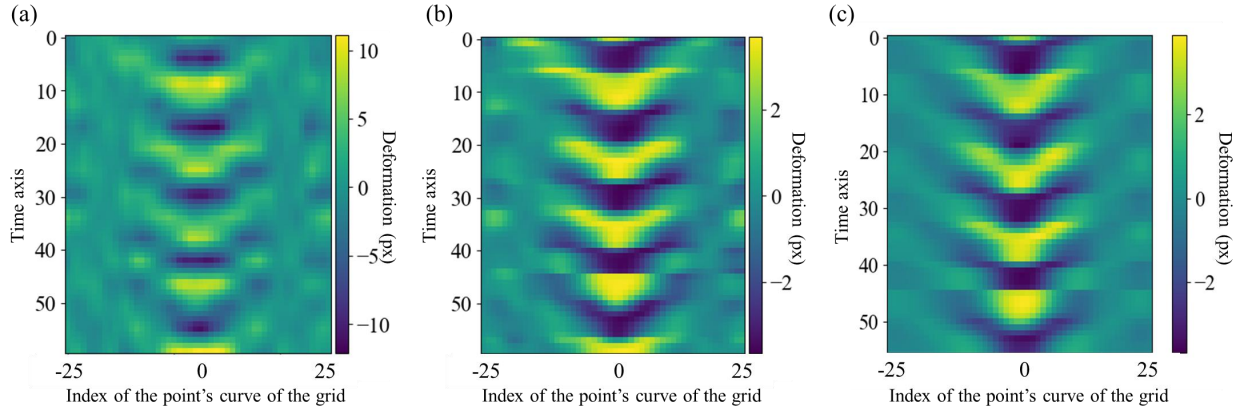


Figure 73 – Successive processing of the map of deformation of Figure 70. Linear correlation (a), correlation of the signal at each time step with a Gaussian function (b) and reconstruction using the two first eigenvectors (c). The horizontal axis corresponds to the number of outer points of the deformation model. The vertical axis corresponds to the time axis.

2.2.3 SIFT for the matching of map of deformation

To analyse the ability to distinguish species from the maps of deformation, we compare them two by two using the SIFT image matching method.

The SIFT method relies on the principle of perspective. In other words, when we look at an object, its importance depends on our field of vision. If our field of vision is small, then we will be more focused on the object than if our field of vision was large. The importance of the object to us will thus be higher for a small field of vision than for a large one. The SIFT algorithm is composed of five main steps that allow the matching of images and that we will briefly introduce based on the work of [Lowe, 2004]. The main steps are the detection of the image's key points, their description and finally their comparison that led to the image matching.

2.2.3.1 Detection of key points

This idea of perspective on which [Lowe, 2004] based its method, is translated via his conception of octaves which correspond to the resizing of image by half of the size of the previous octave's image. At each octave, the image is blurred at different scales (eq. 64). It corresponds to the convolution of a Gaussian kernel, at those different scales, with the input image.

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \qquad \text{(eq. 64)}$$

From all these images generated from the input one, potential key points can be detected. The detection is carried out at each octave, from the Difference of Gaussian (DoG) images [Lowe, 2004]. It corresponds to the difference between two images of a different parameter σ of blurring (Figure 74).

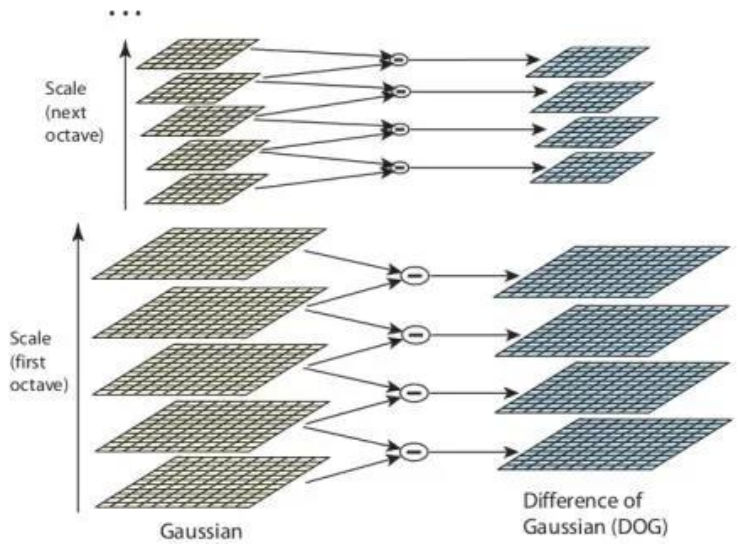


Figure 74 – Explanation of the DoG calculation from the blurred images, from [Lowe, 2004]. An octave corresponds to the image resized and blurred with a different Gaussian kernel.

.

For each DoG image, each pixel is compared with its M immediate neighbours as well as its M+1 neighbours in the previous scale and in the next scale respectively. If the pixel corresponds to a local extremum, meaning if it corresponds to the minimum or maximum intensity of all of the pixels neighbours, then it is considered as one potential key point. As highlighted by [Lowe, 2004], in some images a lot of key points might be detected at this step, [Lowe, 2004] thus proposed to sharpen those detections by removing the key points of low contrast based on pixel intensity and by removing key points lying on edges.

2.2.3.2 Characterisation of key points

After their detection, the key points are characterised. If the previous step allows us to attribute a scale to each key point and thus meet the invariant property of the SIFT algorithm, we still need to meet the second assumption of orientation invariance. An orientation has thus to be

assigned to the key points. A 36-bins histogram of orientation from 0 degree to 360 degrees is calculated from the neighbourhood of the studied pixel. To do so, the magnitude (eq. 67) and direction (eq. 68) of the gradient (eq. 65, eq. 66) of the neighbourhood area are calculated. The magnitude of the gradient of an image highlights how quickly an image changes, while the direction of that gradient helps us to understand in which directions these changes are made [Jacobs, 2005].

$$G_x(r,c) = I_{r,c+1} - I_{r,c-1} \quad \text{(eq. 65)}$$

$$G_y(r,c) = I_{r-1,c} - I_{r-1,c} \quad \text{(eq. 66)}$$

$$Magnitude(\mu) = \sqrt{G_x^2 + G_y^2} \quad \text{(eq. 67)}$$

$$Angle(\theta) = |tan^{-1}(\frac{G_y}{G_x})| \quad \text{(eq. 68)}$$

The histogram calculated thus allows us to summarise all this information. It follows the principle of the Histogram of Oriented Gradients (HoG) [Dalal and Triggs, 2005]. For each pixel of the neighbourhood, the corresponding magnitude of the gradient is added to the histogram bin that matches the gradient direction of that pixel (eq. 69).

$$h(\phi) = \sum_{x \in X} \mu(x) \quad \text{if} \quad \theta(x) \in [\phi - 10, \phi + 10] \quad \text{(eq. 69)}$$

The histogram is then normalised and the peaks that are above 80% are considered. For each of these peaks, a key point is created with the same location and scale and with the corresponding peak's orientation.

2.2.3.3 Key point descriptor

Once the key point is detected and characterised, it is described using a feature vector. The aim is to calculate meaningful information that will later help to compare the key points with other ones. The calculation of the feature vector is carried out on a window of w_l-pixels around the key point. The window is split in n_{block} squared sub-blocks of w_b pixels size (Figure 75) for which a α-bin orientation histogram is generated. The size of the feature vector is thus equal to $\alpha \cdot n_{block}$.

The histogram h_w is calculated from the magnitude and the direction of the sub-block's gradient in the same way as for the orientation assignment in section 2.2.3.2. However, differently from the section 2.2.3.2., the magnitude added in the histogram is weighted using a Gaussian weighting function (eq. 70), with σ equals to half of the descriptor window length. As explained by [Lowe, 2004], the aim is to avoid abrupt changes in the description and to attenuate the importance of gradients that are from the center of the descriptor.

$$h_w(\phi) = \sum_{x \in X} f(\sigma) * \mu(x) \quad \text{if } \theta(x) \in [\phi - 10, \phi + 10] \quad \text{(eq. 70)}$$

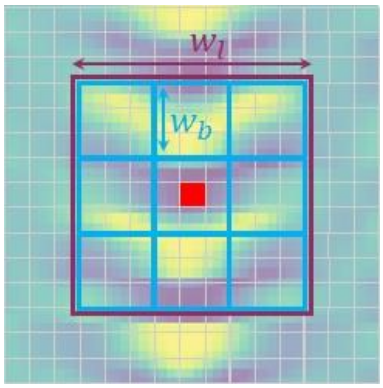


Figure 75 – Explanation of the window and sub-block position from which the histogram is calculated. w_l is the size of the window, in pixels; and w_b is the size of the sub-blocks, in pixels. The example shows the window, split in nine blocks, calculated for the pixel in red.

To ensure that the feature vector is invariant to rotation in case of rotated images, the key point's orientation is subtracted to the values of the gradient direction. In the same way, to make the vector invariant to illumination in case of highly contrasted images, the pixel intensities of the sub-block are normalised and the ones higher than 0.2 are all set to 0.2 [Lowe, 2004]. As explained by [Lowe, 2004], the aim is to reduce the weight of matching the magnitudes for large gradients and rather to accentuate the matching on the distribution of orientations.

2.2.3.4 Key points matching

Key points matching (Figure 76) is carried out using the Nearest-Neighbours approach [Cover and Hart, 1967], meaning that the matching of the key point k_i^I of the image I corresponds to the key point $k_i^{I'}$ of image I' that minimises the Euclidean distance between both feature vectors [Lowe, 2004]. To improve the matching ability and overcome issues of no matching points raised by [Lowe, 2004], he proposed to use a threshold ϵ_r on the ratio of distance r_i between the first and the second closest neighbours (eq. 71).

$$\forall i \in [1, Q], r_i = \frac{d(k_i^I, k_n^{I'})}{d(k_i^I, k_m^{I'})} < \epsilon_r \qquad \text{(eq. 71)}$$

With Q, the number of key points to match; d, the Euclidean distance operator; k_i^I, the point to match; $k_n^{I'}$, the first closest neighbor of the key point to match; and $k_m^{I'}$, the second closest neighbour of the key point to match.

In case of a reliable match, the distance between the first and the second closest neighbours is expected to be higher than in case of a false match.

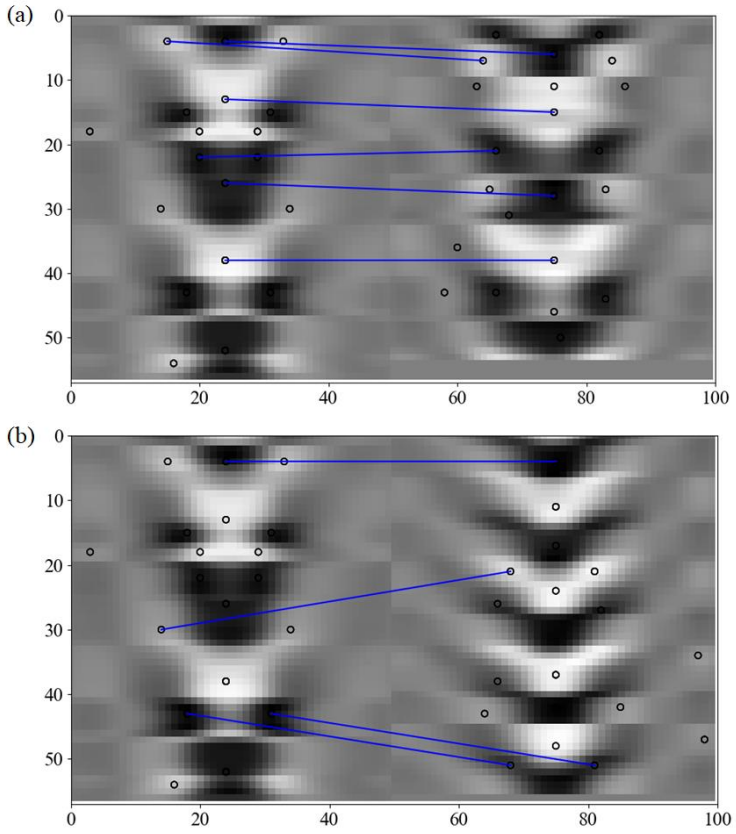


Figure 76 – Examples of key points matching between two maps of deformation. The circles correspond to the key points. The line corresponds to the matching of key points. The matching between two European eels is displayed in (a) and the matching between a European eel and a European catfish is displayed in (b). The vertical axis corresponds to the time axis while the horizontal axis corresponds to the 50 outer points of the deformation grid.

2.2.4 Matching scoring

To classify our maps of deformation, a matching score must be defined. It will allow us to quantify the similarity between the maps of deformation and thus classify the input map of deformation into one class species with the indication of the degree of similarity. We define the index of similarity s (eq. 72) as the maximum of the two ratios R_{IJ} and R_{JI}. R_{IJ} corresponds to the ratio of the total number of matchings between image I and image J, with the minimum number of key points detected on images I and J. Reciprocally, R_{JI} corresponds to the ratio of the total number of matches between image J and image I, with the minimum number of key points detected on images I and J. The matching process is carried out from I to J and then from J to I as a key point i from image I might be the perfect match for the key point j from image J but not reciprocally. By calculating both ratios R_{IJ} and R_{JI} we thus aim to take into account the higher number of matchings. This calculation is made only if a minimum of three matchings, threshold empirically chosen, is generated by the SIFT matching method.

$$s = max(\frac{M_{IJ}}{min(K_I, K_J)}, \frac{M_{JI}}{min(K_I, K_J)}) \quad \text{(eq. 72)}$$

With, M_{IJ}, the number of matching combinations for the comparison of image I with image J; M_{JI}, the number of matching combinations or the comparison of image J with image I, and, K_I and K_J, the number of key points detected respectively on image I and image J.

The index of similarity is considered in the analysis if its value is above 40%, below we considered that the index is too low to describe a good image matching.

3 Preliminary results

Preliminary results of the pipeline are presented in this section. Our objective is to explore the ability to distinguish between species or groups of species from the maps of deformation. Subsets of the MZC-ARIS and SEL-ARIS datasets, all recorded with ARIS cameras, are used as a dataset of test. A total of 31 fish passages is analysed: 10 European eels, 10 Atlantic salmons, 5 European catfish, 5 fish from the CYP group and 1 from the PRED group. The maps of deformation are displayed and visually inspected in regards to the trajectory of the fish passages. Prior to calculating similarity scores between the maps of deformation, an exploratory investigation is then conducted in order to analyse the distribution of the deformation for each point of the curve's grid over time. We finally analyse more in detail those maps of deformation by looking at deformation patterns and their similarities between species and groups of species.

3.1 Visual representation of the body deformation

Figure 77 displays the fish body deformation of three individuals of our dataset of test: one European eel (Figure 77.a), one European catfish (Figure 77.b) and one Atlantic salmon (Figure 77.c). Three successive vignettes centred on the fish are displayed with the deformation grid that follows the fish body shape.

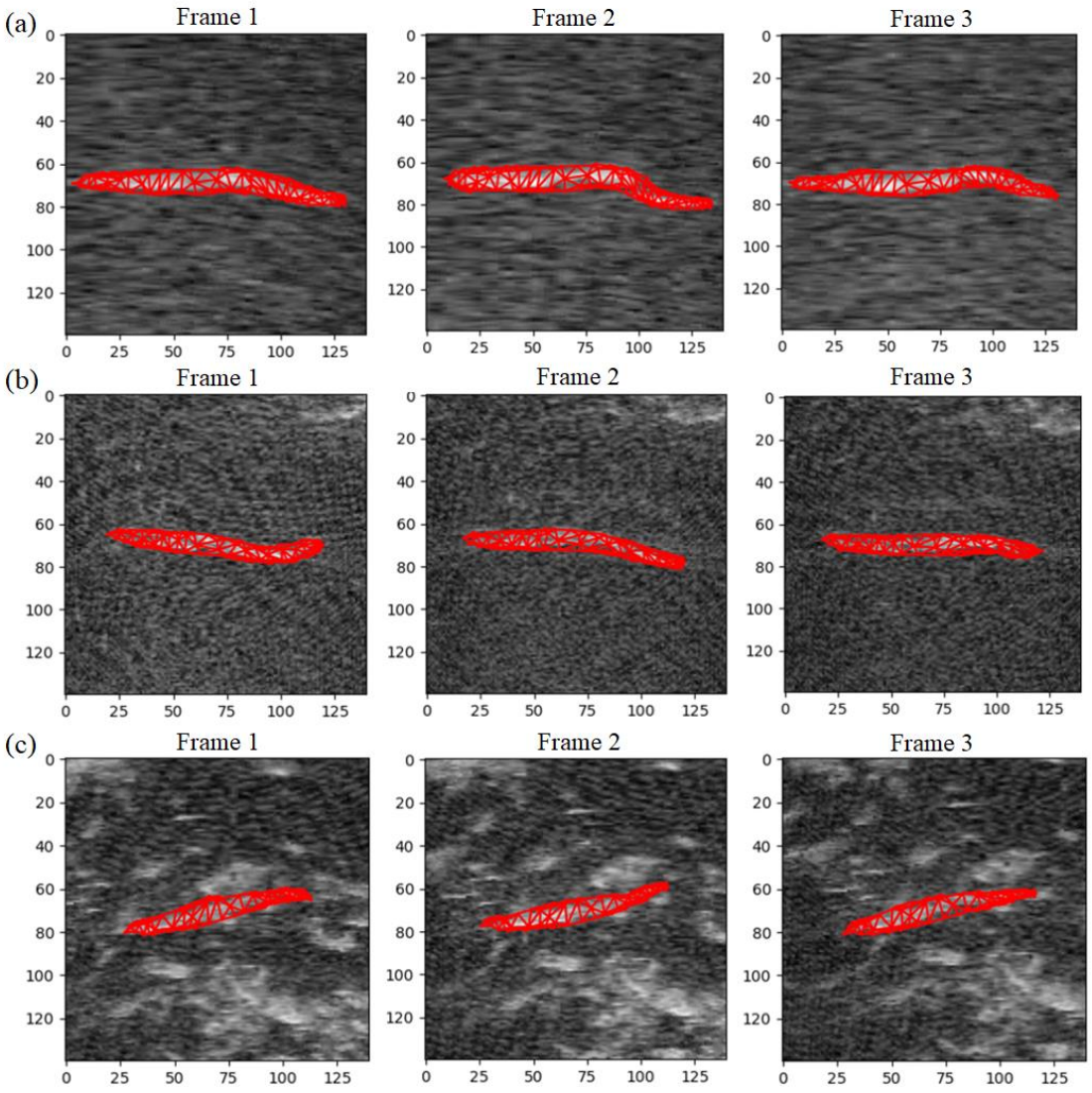


Figure 77 – Examples on three successive frames of the fish body deformation for one European eel (a), one European catfish (b) and one Atlantic salmon (c), of similar size. The deformation grid is displayed in red.

3.2 Maps of deformation for preliminary analysis

Post-processed maps of deformation of the fish from the dataset of test are displayed in Figures 78, 79, 80, 81 and 82. The sizes of the maps of deformation vary from one to another as the duration of detected passages differ from one fish to another.

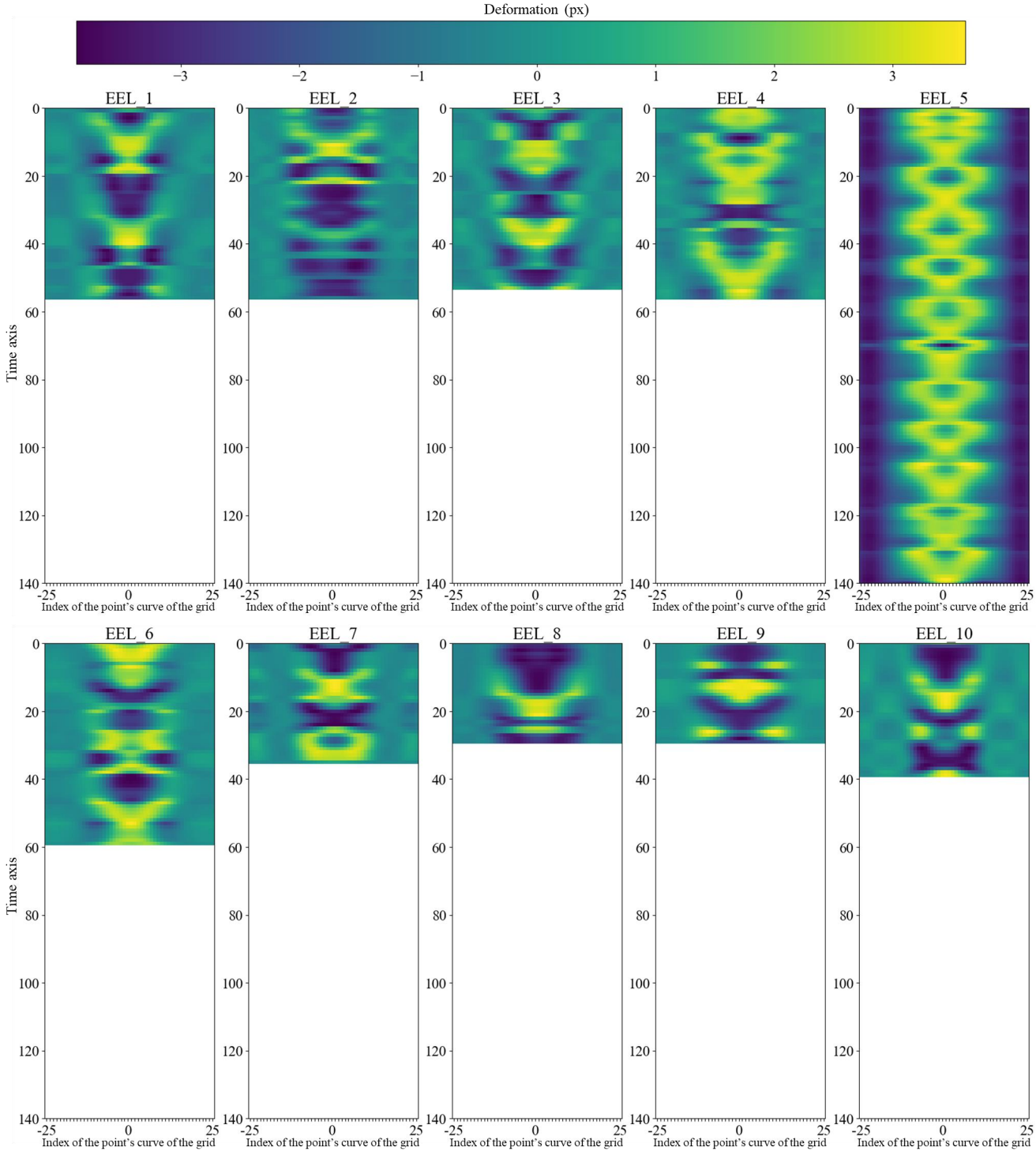


Figure 78 – Post-processed maps of deformation of the European eels passages (N = 10). The first half on the left side corresponds to the points of the upper side of the fish silhouette, from the head to the tail. The second half corresponds to the points of the lower side of the fish silhouette, from the tail to the head. The colour bar corresponds to the range of the deformation values, in pixel.

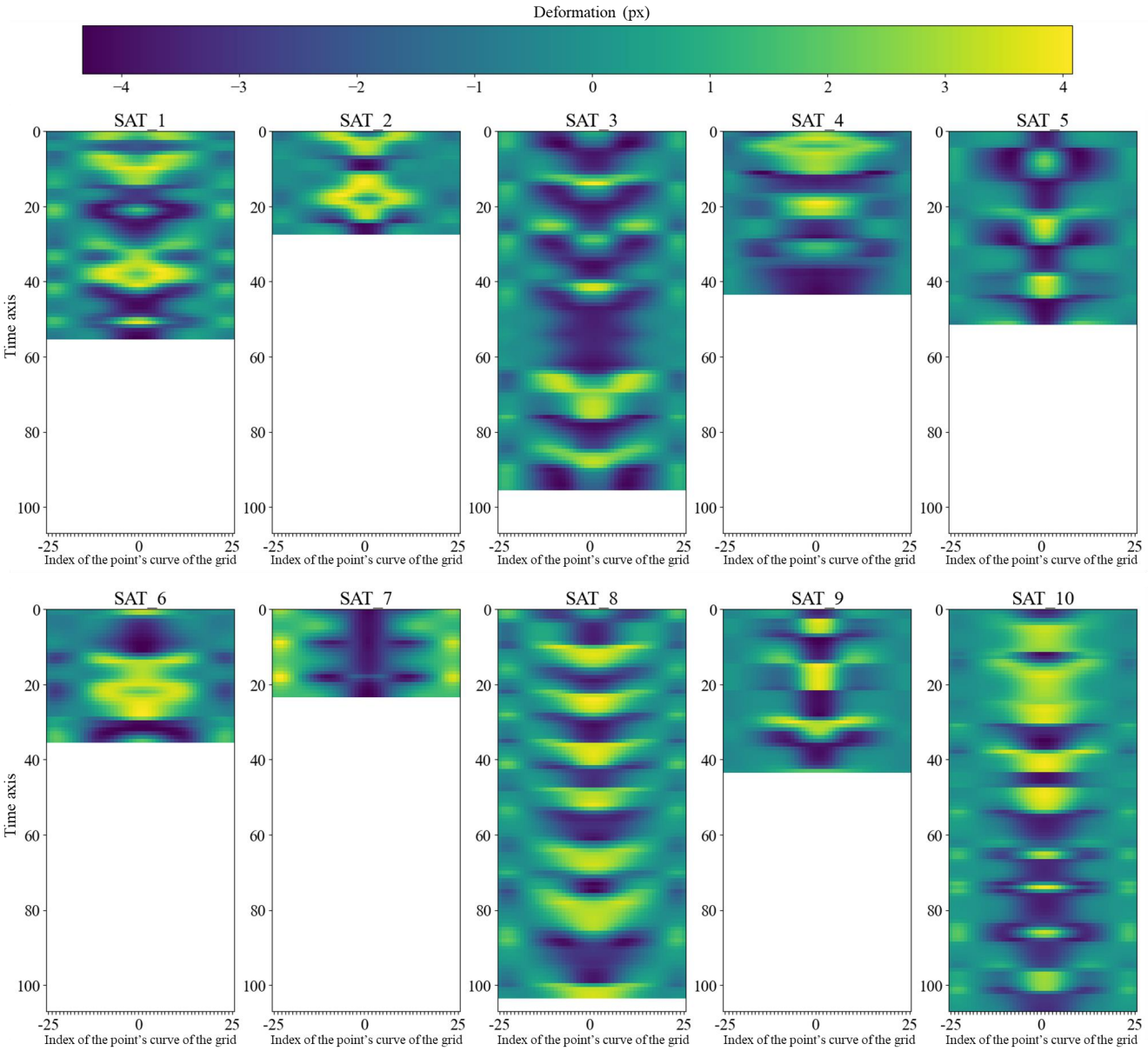


Figure 79 – Post-processed maps of deformation of the Atlantic salmons passages (N= 10). The first half on the left side corresponds to the points of the upper side of the fish silhouette, from the head to the tail. The second half corresponds to the points of the lower side of the fish silhouette, from the tail to the head. The colour bar corresponds to the range of the deformation values, in pixel.

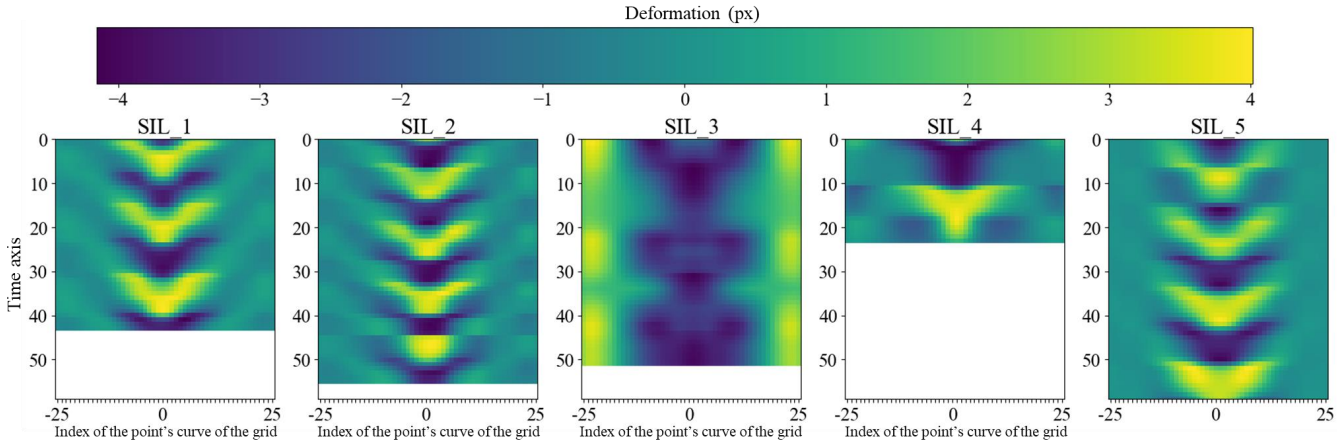


Figure 80 – Post-processed maps of deformation of the European catfish passages (N = 5). The first half on the left side corresponds to the points of the upper side of the fish silhouette, from the head to the tail. The second half corresponds to the points of the lower side of the fish silhouette, from the tail to the head. The colour bar corresponds to the range of the deformation values, in pixel.

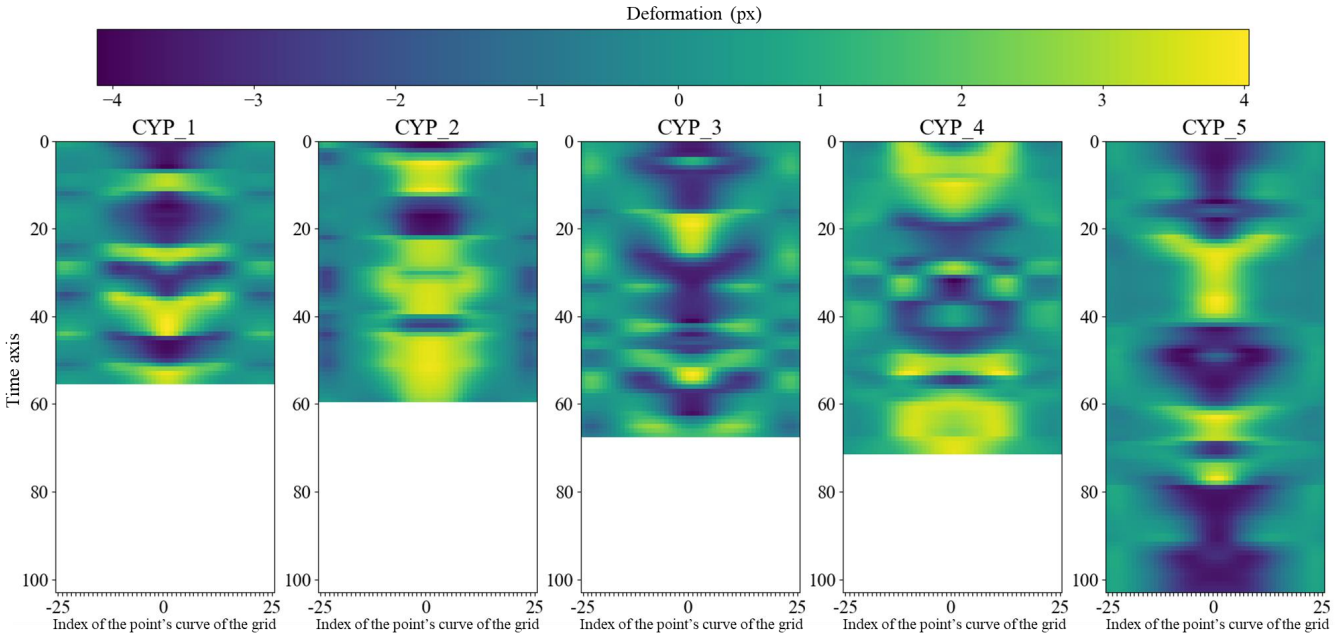


Figure 81 – Post-processed maps of deformation of five passages of a fish from the CYP groups of species. The first half on the left side corresponds to the points of the upper side of the fish silhouette, from the head to the tail. The second half corresponds to the points of the lower side of the fish silhouette, from the tail to the head. The colour bar corresponds to the range of the deformation values, in pixel.

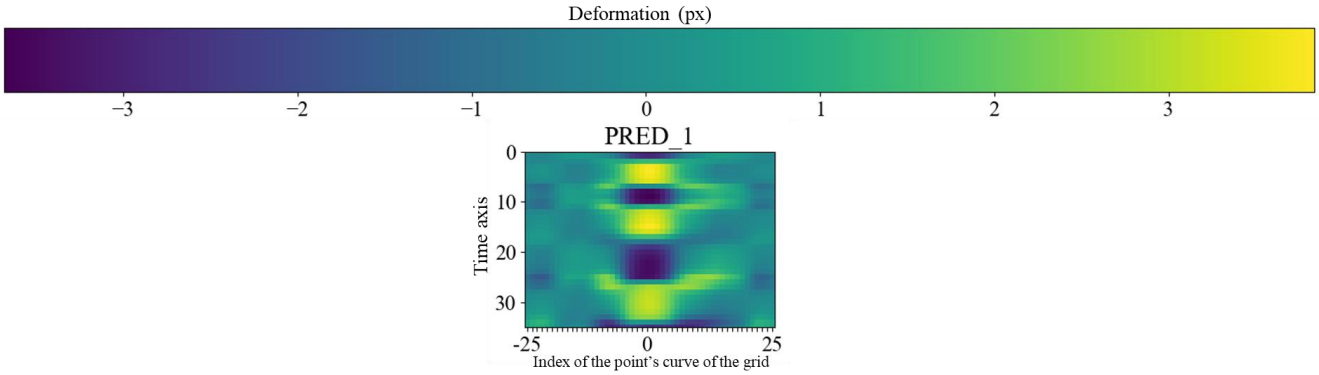


Figure 82 – Post-processed maps of deformation of the passage of the fish from the PRED group of species (N = 1). The first half on the left side corresponds to the points of the upper side of the fish silhouette, from the head to the tail. The second half corresponds to the points of the lower side of the fish silhouette, from the tail to the head. The colour bar corresponds to the range of the deformation values, in pixel.

The maps of deformation of European eels seem at first different from each other (Figure 78). However, common patterns can be observed between some of the individuals. On the one hand, EEL1, EEL2 and EEL6, on the other hand EEL2, EEL4, EEL7 and EEL10, and finally, EEL1 and EEL3, have rather similar deformation. However, patterns can be represented more or less stretched depending on the individuals which may rely on the velocity of the body deformation and therefore on the fish velocity. All of these individuals have crossed straight the AC FOV with similar trajectories except for EEL5. Hence, unlike the other eels in the dataset, EEL5 swims against the current and perpendicular to it (Figure 83.a) which may explain that its passage provides a completely different map of deformation compared to the other of the group.

Similar conclusion can be drawn for the Atlantic salmon passages (Figure 79). On the one hand, SAT3, SAT5 and SAT9, and on the other hand, SAT1, SAT2, SAT4, SAT6 and SAT10 display similar patterns even if SAT4 and SAT10 appear as compressed. All of their trajectories are quite heterogeneous, especially SAT10 swims following a hazardous trajectory, that might explain the high variability observed on those maps (Figure 83.b). However, two individuals remain completely different from the other of the group: SAT7 and SAT8. The map of deformation of SAT8 even seems to be similar to the ones of SIL4 and SIL5 although the patterns of SAT8 are wider, the body deformation being not limited to the caudal fin.

Compared to the European eels and the Atlantic salmons, the European catfish (Figure 80) is one of the species with the least visual intra-species variability on the maps of deformation. Hence, the V-shaped pattern is recognisable on 4 out the 5 individuals in the dataset of test with an alternation of the deformation direction. The observed deformation also appears to be periodic, although the passage of SIL4 only allows the observation of one complete deformation cycle. The only exception to these observations is SIL3. Analysis of its trajectory (Figure 83.c)

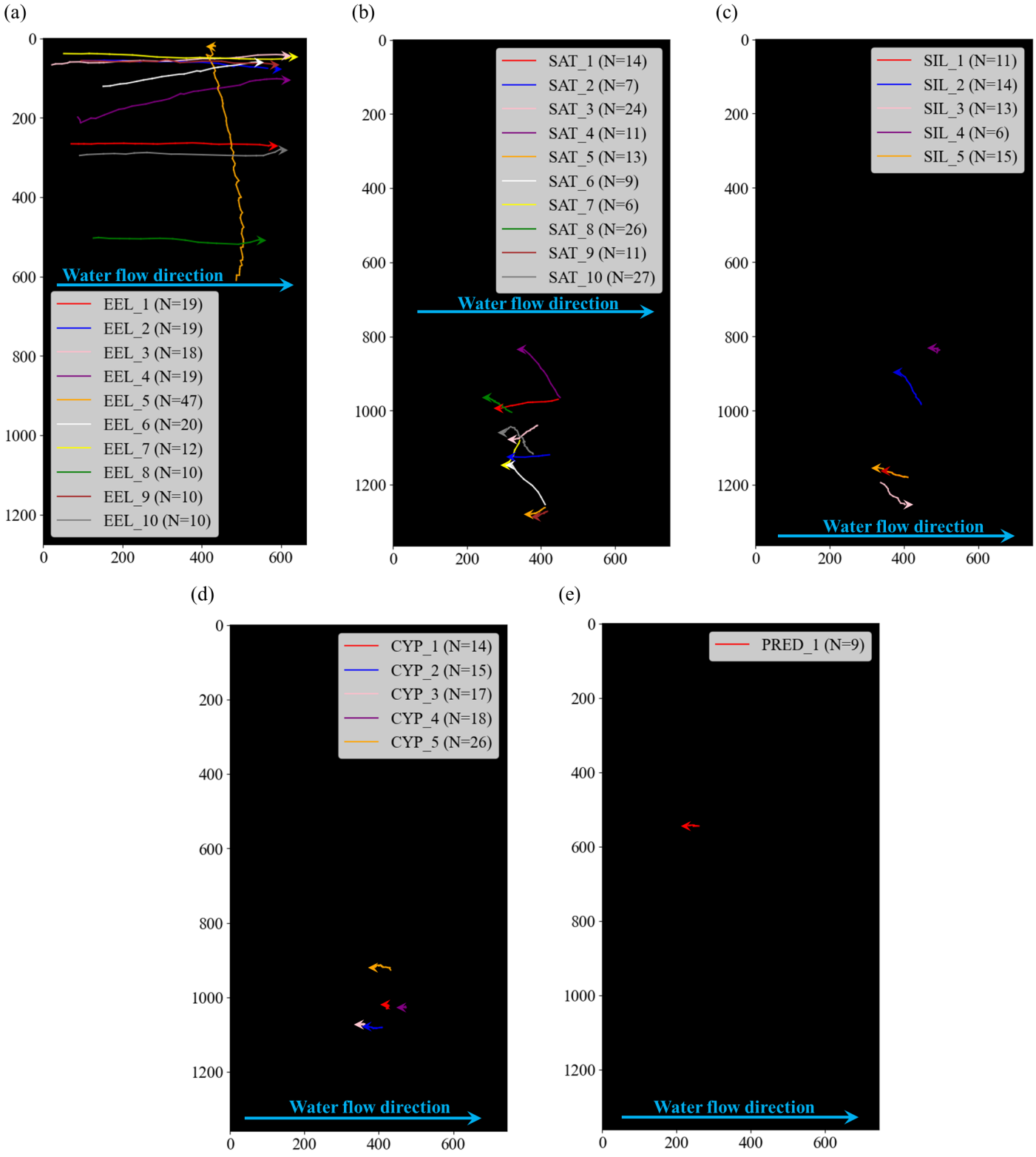


Figure 83 – Trajectories in the AC FOV of ten European eels (a), ten Atlantic salmons (b), five European catfish (c), five fish from the CYP group of species (d) and one fish from the PRED group of species (e).

highlights a random movement in the direction of the current. The video of the passage shows that SIL3 seems to be carried by the current while the other four European catfish actively swim against it.

Concerning the CYP group of fish (Figure 81), the maps of deformation highlight high intra group variability similar to the SAT group. The trajectories are heterogeneous as well, with individuals travelling very short distances compared to the EEL group, for a similar number of detections. CYP3 and CYP5 might even display similar patterns as those observed in some Atlantic salmons.

Besides, this caudal beat is observed in the fish of the PRED group. However, only a one off observation can be provided, as no comparison with another individual of the same group can be made.

3.3 Descriptive statistic for maps of deformation analysis

The deformation values of each point of the curve along the detection frames are displayed per each group of species, as box plots (Figure 84). From these box plots, it can be noticed that the range of deformation is similar for all the four groups of species. Besides, by reading the extreme points of the deformation distribution, box plots highlight that European eels and European catfish are moving their head less compared to the fish from the SAT, CYP and PRED groups of species. Besides, the fish from those last groups are deforming their whole body when swimming, high deformation is observed between the points number -10 and 10. In contrast, the deformation of the European eels and European catfish bodies seem to be more centred on the last third of the body, around the caudal fin.

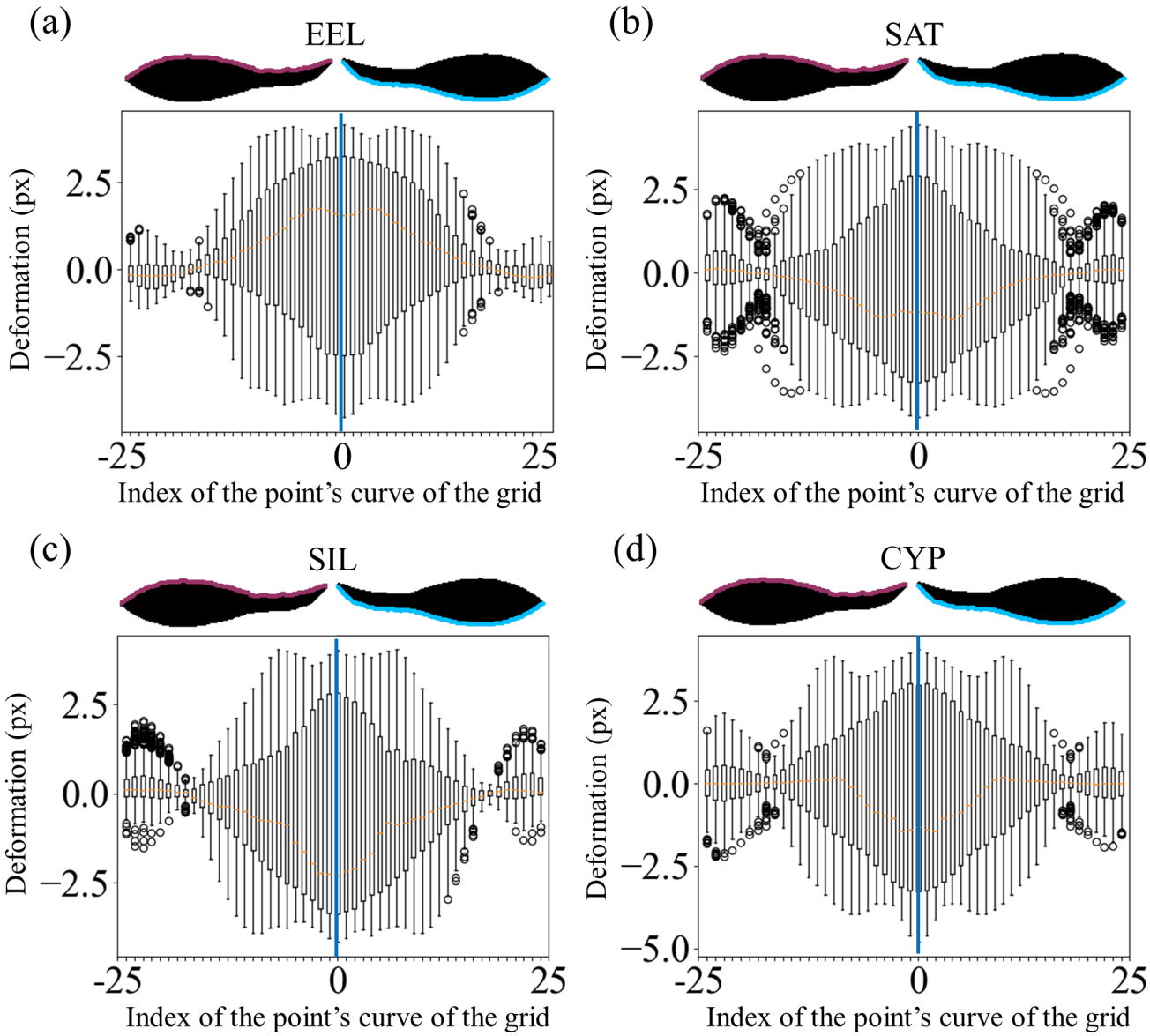


Figure 84 – Box plots of the post-processed maps of deformation of the EEL (a), SAT (b), SIL (c) and CYP (d) groups of species. The box extremities correspond to the minimum, without the ouliers, the first quartile, the third quartile and the maximum, without the ouliers. The median of each box is represented as an orange horizontal line and the outliers are represented as circle.

3.4 Automatic analysis of the maps of deformation: analysis of the similarities

In order to give indexes of similarity between all of the 31 fish of the dataset of test, their maps of deformation are analysed two by two using the SIFT method. It allows us to focus on the specific pattern present on each map of deformation, highlighted by our visual analyses. Figure 85 displays the resulting scoring matrix. These indexes of similarity give an indication of the capacity to distinguish species from the maps of deformation.

The visual analysis of Figure 85 highlights intra-group similarities between individuals for the species and groups of species EEL, SIL and CYP. Regarding the European eels, eight out of the 10 individuals of the group are similar to at least one of their group mates. Only EEL2 and EEL5 are not considered similar ($s > 40\%$, section 2.2.4) to any other individuals, even from another group. Concerning the European catfish, 3 of the individuals are considered similar to each other while SIL3 and SIL5 are not similar to any of the 31 individuals in the dataset. Finally, 4 out of the 5 individuals of the CYP group of species show similarities with the other individuals of the group, CYP4 not matching any of them.

Although numerous intra-group similarities are observed in these three groups, this is less the case for the SAT group. Out of the ten individuals studied, only 3 show similarity with another individual of the group, the 7 other individuals do not match any of them nor individuals of the other groups. However, inter group similarities are also shown for the European eels, especially between EEL7, EEL8 and EEL9, with SAT5, SAT9, SAT10, SIL2, SIL5, CYP2 and CYP3. Conversely, individuals from the CYP species group, in addition to showing similarities with the European eel, also show correspondences with the Atlantic salmons as it is the case for CYP2. The only groups that show no similarities are the European catfish and the CYP group.

However, the similarities listed below reach rather low similarity indexes, mostly between 45 and 60%.

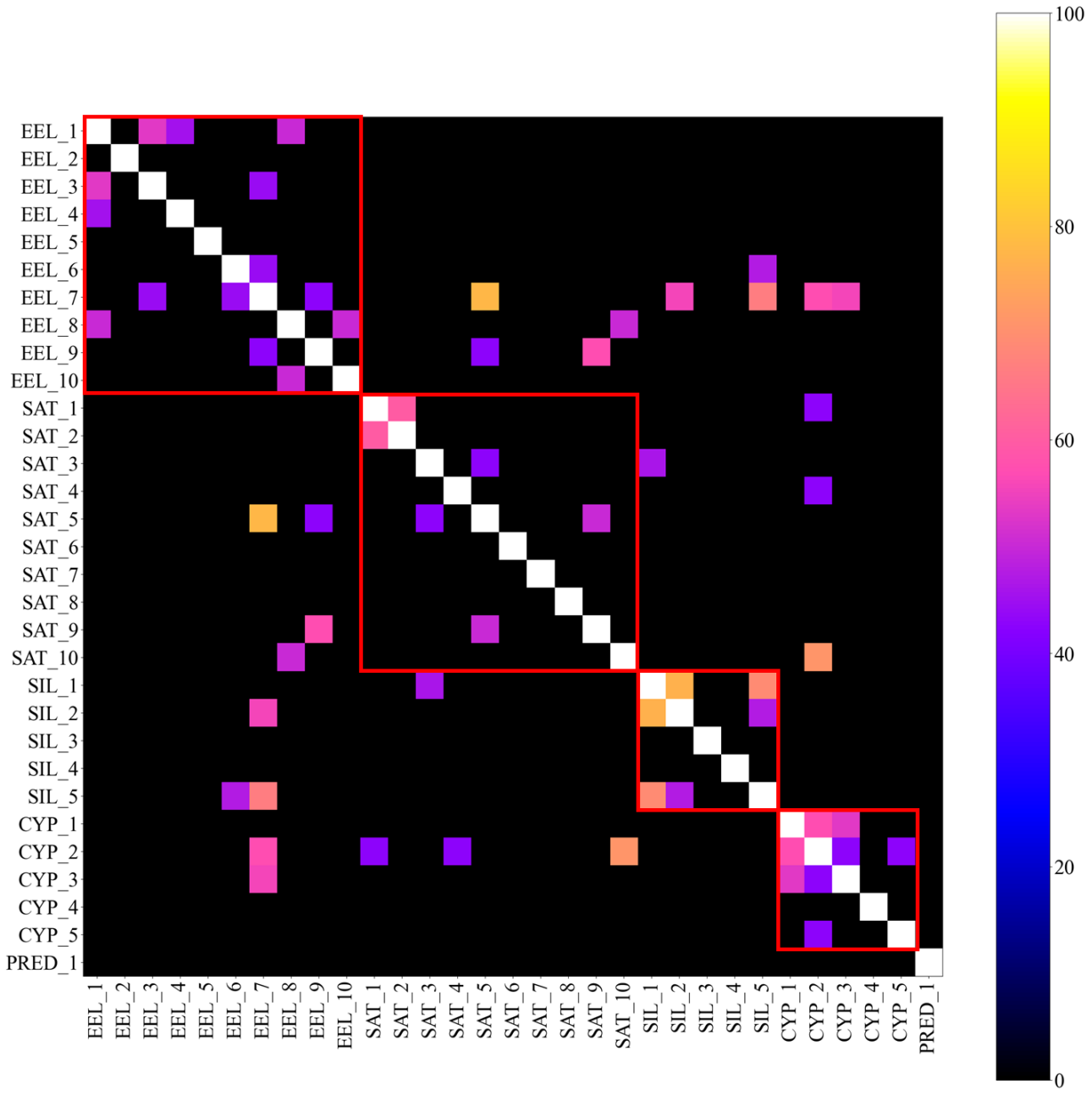


Figure 85 – Scoring matrix obtained by comparing each map of deformation of the 31 individuals of the dataset of test, using the SIFT algorithm and our index of similarity s. The scoring matrix threshold is equal to 40%, only the indexes of similarity whose values are above this threshold are displayed in colour. The colours depend on the value of the index of similarity s, from 0 to 100%. The lighter the colour is, the more similar the two individuals are. The red squares correspond to similar species or groups of species.

4 Discussion

This chapter introduced the method we developed for the characterisation of the fish body deformation along its movement in the AC FOV. The aim is to (a) follow up the fish body deformation from one frame to another, (b) quantify this deformation at each time step, and finally (c) evaluate the species discrimination ability of those calculated deformation.

The gradient domain mesh deformation model [Weng et al., 2006] we used perform well on our videos allowing us to closely follow-up the fish body contours at each of its detection in the camera FOV. However, it must be noticed that the binary image of the fish body given as an input of the model will largely influence the deformation results. In case of noise or occlusion in the fish body image, the model will deform according to these artefacts rendering an incorrect mesh deformation of the body. The resulting calculation of the deformation might thus also be wrongly impacted. It is not the only factor influencing the deformation quantification. As previously mentioned, models of deformation are not designed to handle big translations or orientations. Our handling of those big changes through rotation and translation, might also influence the final deformation as the choice of another way to match two successive fish bodies will change their mesh positions towards each other and thus the deformation calculated. All of these potential sources of errors have to be brought to the attention of the operator to understand potential bias. The binary image of the fish, used to calculate the body deformation, therefore highly influences the exactitude of the process and highlights the importance of each previous step of the acoustic videos analysis.

As calculated through the models of deformation, the maps of deformation are hardly readable and thus hardly comparable. Their post-processing was therefore an important step before retrieving any valuable information. Studies on swimming mode proposed in the literature are carried out from optic cameras known for their frame rate of 24 frames per second [Eguiraun et al., 2014], enabling a good follow-up of the fish deformation along their movement. With our AC whose frames rates vary from 5 to 8 frames per second, we can catch the deformations but in a jerky way. Interpolation of our maps of deformation along the time axis therefore smooth the transitions and highlight the deformation process the fish body is submitted to. We choose to use a linear interpolation in our pipeline but other methods exist. It is an arbitrary and non-trivial choice as it influences the maps of deformation we are then expecting to analyse. The two following steps of filtering were carried out in order to clean those maps by focusing on their main characteristics in order to simplify the comparison process. The parameters σ of the Gaussian which are correlated with the map of correlation, have been qualitatively chosen to avoid smoothing too much the signal and taking the risk of attenuating important peaks. The last step of the process using Singular Value Decomposition, finally highlights the main characteristic of the deformation. However, as for the interpolation,

both choices of the σ parameter and of the eigenvectors used influence the final information analysed and may bring some bias. However, these processes also allow us to retrieve readable information that we are able to analyse regarding species identification.

The box plots based on the deformations of the different species show similarities between the species with anguilliform swimming mode, EEL and SIL, as well as between the species with subcarangiform swimming mode, SAT and CYP, which is consistent with the classification of swimming patterns proposed by [Breder and Society, 1926, Lindsey, 1978].

However, a more detailed visual analysis of the maps of deformation highlights a high variability of the maps of deformation from one individual of a group to another individual of the same group. This variability influences the cross-comparison results calculated via the pairwise matching of maps of deformation and leads to matchings between individuals of different groups, for example between European eels and European catfish. The anguilliform swimming mode of both species can however explain these results. Same conclusion can be drawn regarding the occasional similarities between fish of the SAT and PRED groups that both have a subcarangiform swimming mode. However, some similarities are less consistent even after visual inspection of the maps of deformation, for example for SAT5 and EEL7, which suggests that another method of calculating similarity could be used rather than the image matching method via SIFT. A template searching method aiming to match small parts of the input image with a template image, could be considered. The template image would correspond to the characteristic patterns of each species or group of species, could for example be considered. To avoid any operator bias in pattern selection, it will be essential to automatically identify them prior to the search.

However, the results are overall encouraging with similarities observed between individuals from the same groups of species, for example for the EEL, SIL and CYP groups. Some individuals remain unmatched with their group mates, and a visual analysis may explain these results. For SIL4, although the visual analysis shows a strong similarity of its map of deformation with SIL1, SIL2 and SIL5, it does not emerge with the matching method because of a too small number (3) of key points detected.

The trajectories of EEL5 and SIL3 largely differ from the other individuals included in their groups, causing differences in their maps of deformation, and consequently explaining that they do not match with other individuals of their respective group. While the other European catfish move upstream, SIL3 swims downstream, in the direction of the current, which may explain the difference in its swimming pattern and therefore the difference in its map of deformation. Similar conclusion can be drawn for EEL5 which, unlike the other individuals of its group, swims upstream. Furthermore, the movement of EEL5, parallel to the camera beams, may

also influence the retrieved image of its body on the acoustic videos and thus the deformation tracking leading to the calculated map of deformation. Hence, depending on the orientation of the fish, a part of the body may not be reached by the acoustic waves and thus not appear on the acoustic images. If the image restoration step allows us in this case to recover a continuous image of the fish, this may still include a bias in the deformation. Considering the fish's orientation towards the camera beams could allow, as for the trajectory, a better understanding of the maps of deformation and should thus be further investigated. The analysis of these maps of deformation allows us to highlight different swimming behaviours and even changes in behaviour during the fish movement.

These observations therefore suggest that external factors may influence the maps of deformation and should therefore be further considered in their analysis. Water flow, direction, trajectory and fish speed are all influencing the swimming behaviour of fish and therefore the deformation of their bodies. To allow a more robust analysis of the calculated maps of deformation, it is necessary to include more fish passages with different configurations of these factors allowing to extract the characteristic patterns of their deformation in each configuration. A laboratory study, in a large swimming channel, would be a first approach to study the passages of different species. Hence, to further develop our method on body deformation, it is essential to increase the number of fish passages studied, especially for the SIL, CYP and PRED groups, which are the least represented in this dataset.

Thanks to these preliminary results, this chapter illustrates the high variability of body deformation existing between individuals, even in the same species, making it difficult to obtain a specific pattern. Furthermore, numerous factors may influence the fish body deformation, and its detection by the method we developed. However, the results remain encouraging by highlighting similarities between fish from the same species swimming in a similar way into the AC FOV, making of fish body deformation study a promising characteristic to measure in order to distinguish species. The information brought by the maps of deformation would also allow to study the fish behaviour during its passage in the AC FOV.

Discussion: fish detection and species identification on acoustic camera imaging

Contents

On ne saurait être l'homme de sa spécialité,
que si l'on est aussi sa victime.
Friedrich Nietzsche

Acoustic cameras are less sensitive than video cameras to environmental conditions such as water turbidity allowing continuous recording of underwater life. They are increasingly used for fish movements monitoring and have a lot of potential for aquatic biodiversity monitoring. However, such devices create large volume of data whose manual analysis requires a large operator cost. Because of the time-consuming process of visualisation, operators often only review a few samples of the whole acoustic videos datasets, and consequently only collect a small part of the biological information recorded by AC. Real-time analysis of this data is therefore hardly possible. Besides, because of the limited information displayed by acoustic imaging, the results of the video review by an operator can be highly affected by his or her experience, which may add bias to the fish detection and species identification. Automatic analysis of acoustic videos therefore appears essential to overcome these constraints and allows the full exploitation of acoustic imaging devices.

My PhD work explores approaches that have been little or never investigated in previous studies, in order to automate the analysis of acoustic videos. It has challenging objectives that meet AC users expectations to complete analysis of their data: I have developed tools based on computer vision methods that propose target detection and species identification, which can be transposed to a large panel of acoustic cameras, independently of the monitoring sites, while being compatible with real-time data processing. This chapter highlights major results of my work, and presents advantages, current limits and future developments needed (Figure 86).

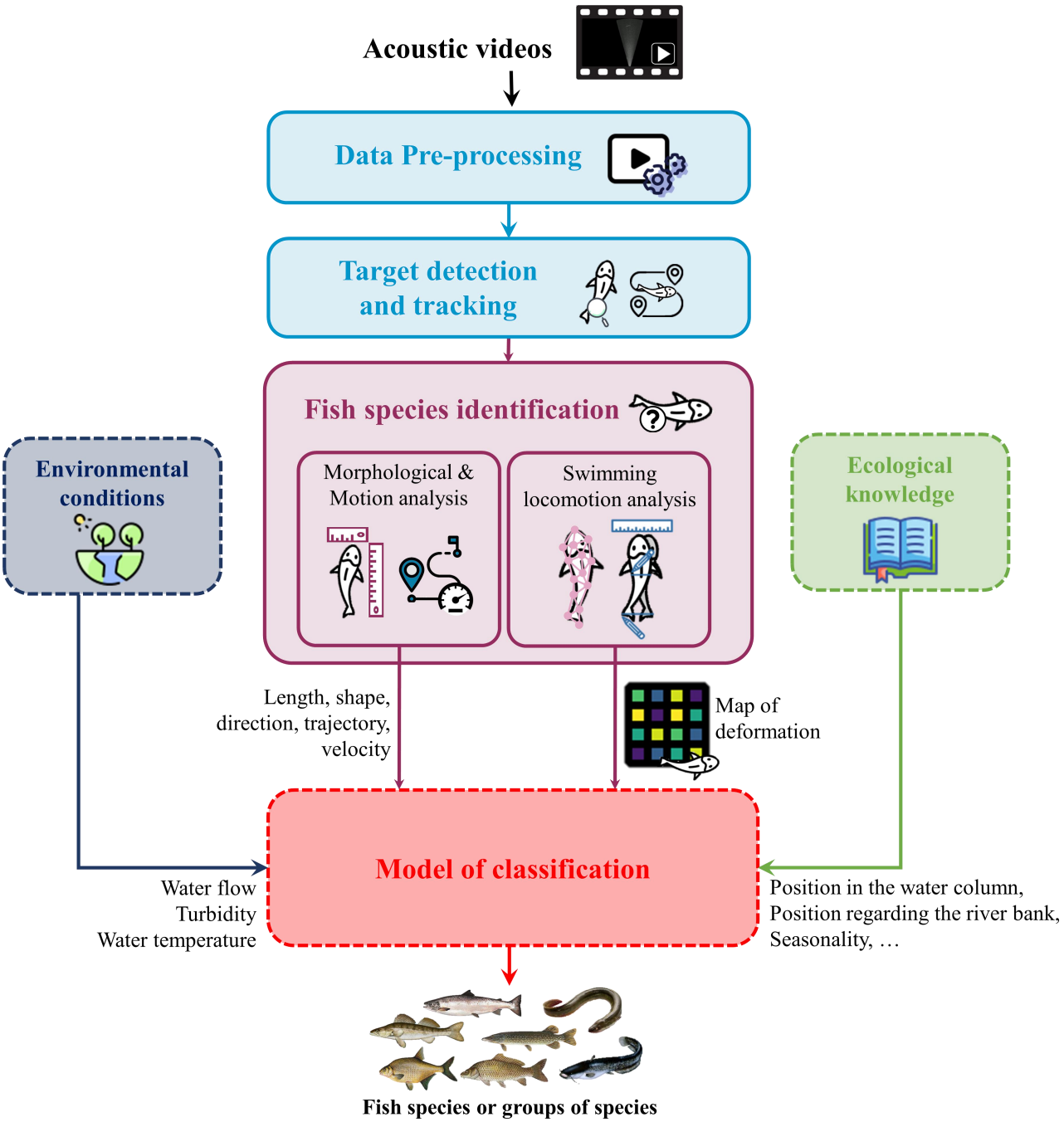


Figure 86 – Proposed pipeline with the different steps of the process displayed. The steps introduced in this work (light blue and pink) and the ones that still need to be developed (in dotted line).

1 Studying living organisms in shallow waters: the challenge of acoustic data

Identification of fish species on videos is a field of study widely investigated in the literature. These videos are mainly recorded by optical cameras allowing a high level of detail and therefore good performances on the automatic recognition of species, but they are only performing at short distance. Hence, optical cameras record numerous fish morphology characteristics which are useful clues to distinguish species: overall shape, number and position of the fins, but also the fish appearance, their body colours and the patterns of their scales (Figure 87.a). Without such details, the identification process is much more complex (Figure 87.b), explaining why the identification of fish species on acoustic imaging is a challenge (Figure 87.c).

Figure 87 – Image of an Atlantic salmon captured by an optical camera at a fish way location © Fédération de pêche de la Somme (a) and the corresponding overall silhouette in black (b). Echo of an Atlantic salmon recorded on an ARIS 1800 image (c).

Analysis of living organisms such as fish with AC provides high intra-individual variability: the same individual may generate dissimilar acoustic images along its trajectory, which is another source of challenge for the fish automatic identification. Moreover, in an uncontrolled environment and unlike laboratory studies, environmental conditions, such as debris, vegetation or obstacles, add variability to the data, particularly in terms of quality. Depending on the water flow or turbidity of the river as well as on the aquatic vegetation development, the analysis of acoustic videos will be more or less challenging. Hence, the quality of the acoustic images is a key element to achieve correct automatic analysis. This is why a pre-processing step focusing on image reconstruction and data pre-analysis, i.e., sorting images that can provide reliable information, is required.

The acoustic videos are reconstructed through a 2D representation of the echoes insonifying a water volume and reflected to the AC. These reconstructions can be the source of broken targets, i.e., the image of a target is fragmented into several regions of pixels of similar intensities (Figure 88.a) instead of being represented by a continuous region (Figure 88.b). The image restoration step is therefore essential to correctly analyse target's detections and retrieve

reliable morphological and body deformation features.

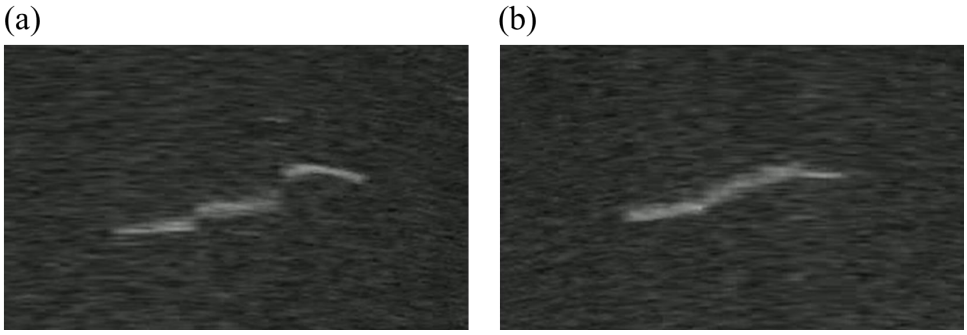


Figure 88 – Broken echoes (a) and correct echoes (b) of a European eel recorded on two ARIS images.

Besides, depending on the position and orientation of the target towards the camera beams, its acoustic image will differ. Hence, if a target swims perpendicularly to the beams from one side of the FOV to the other (Figure 89.a), its whole body reflects the echoes and its whole body appears on the acoustic image (Figure 89.b). Conversely, if standing along the beams height (Figure 89.c) or if it arrives facing the beams (Figure 89.e) or diagonally to the beams (Figure 89.g), only a part of its body is likely to appear on the image (Figure 89.d, f and h). Morphological characteristics may then be distorted and must be carefully integrated in the identification process. A pre-analysis should be carried out to only select the most relevant images from a target passage. For example, morphological features calculated on a target, when its body has not yet fully entered the AC FOV, are not representative of the target and may therefore bias the analysis. It influences the morphological analysis but also the analysis of the body deformation because the grid will have to deform more to fit the whole target's body at the next detections.

In the pipeline I propose, a preliminary selection is made on the targets detections. The sorting is carried out according to the target successive lengths to keep only the longest detections. Target's orientations in the insonified volume should provide an indication on the portion of the body that reflects echoes but are not available from AC data. The selection of only the most consistent echoes reflected by the target during its detection highlights the usefulness of exploiting the entire passage of the target.

The quality of the information also depends on the acquisition settings and on the properties of the monitoring site. Acquisition in open water configuration, i.e., when none of the AC beams encounter the riverbed, gives images with only objects passages (Figure 90.a) and with no echoes from submerged concrete structure like the river's bottom (Figure 90.b): fish passages are consequently easier to distinguish from the background. Besides, the greater the

number of pixels defining an object, the more information the automatic pipeline can collect and therefore the more reliable the identification is. Although some camera and acquisition settings enable to collect acoustic images of good quality (open water, small window length), they do not always match the constraints inherent to the monitoring site (low water depths, high window length required to maximise the AC coverage).

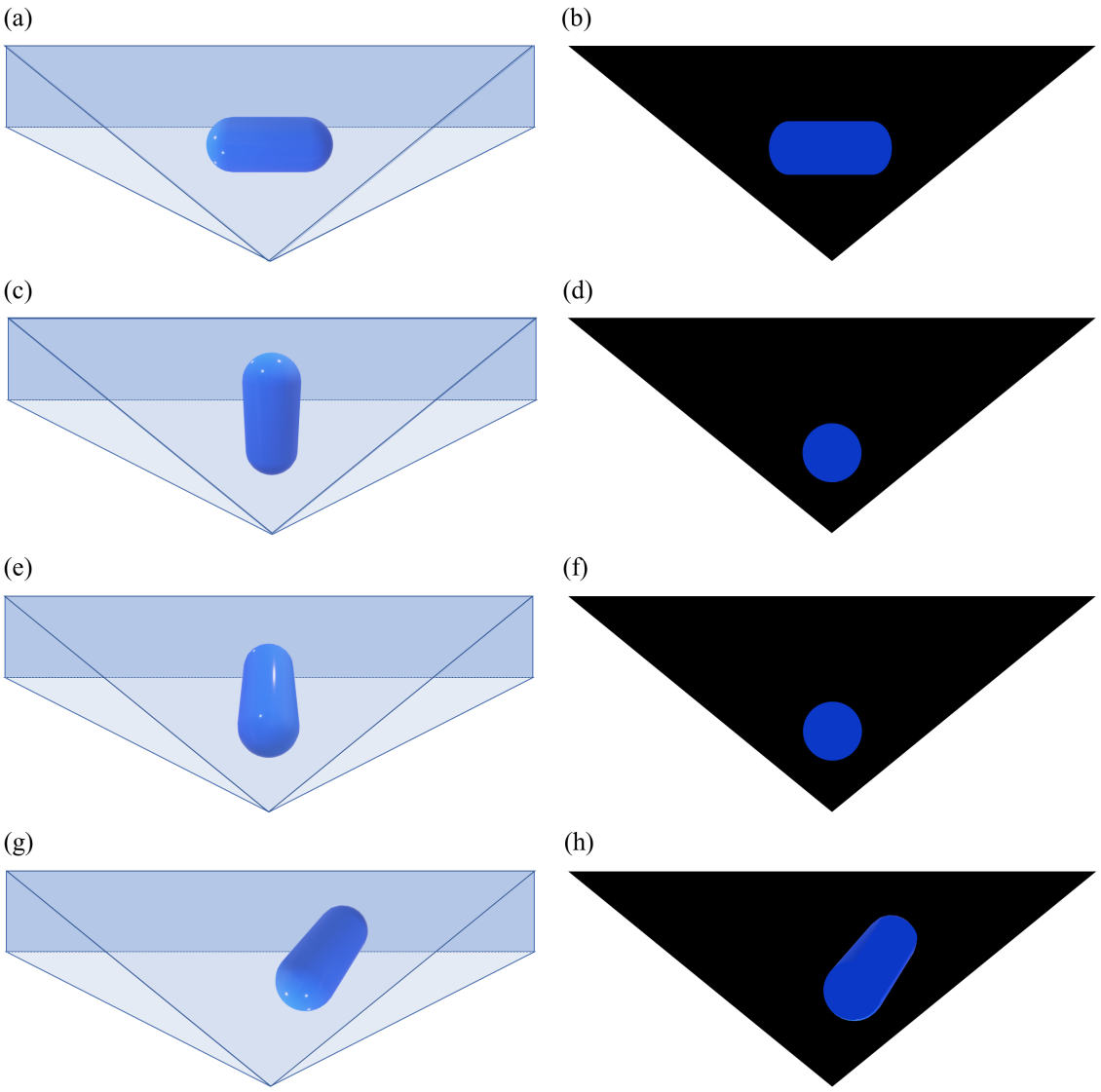


Figure 89 – Schemes of a target passing in the AC FOV (first column) and the corresponding acoustic images (second column) depending on the position of the target: passing horizontally in the AC FOV and perpendicularly to the AC (a), facing vertically the AC (c), facing horizontally the AC (e) and passing diagonally to the AC beams (g).

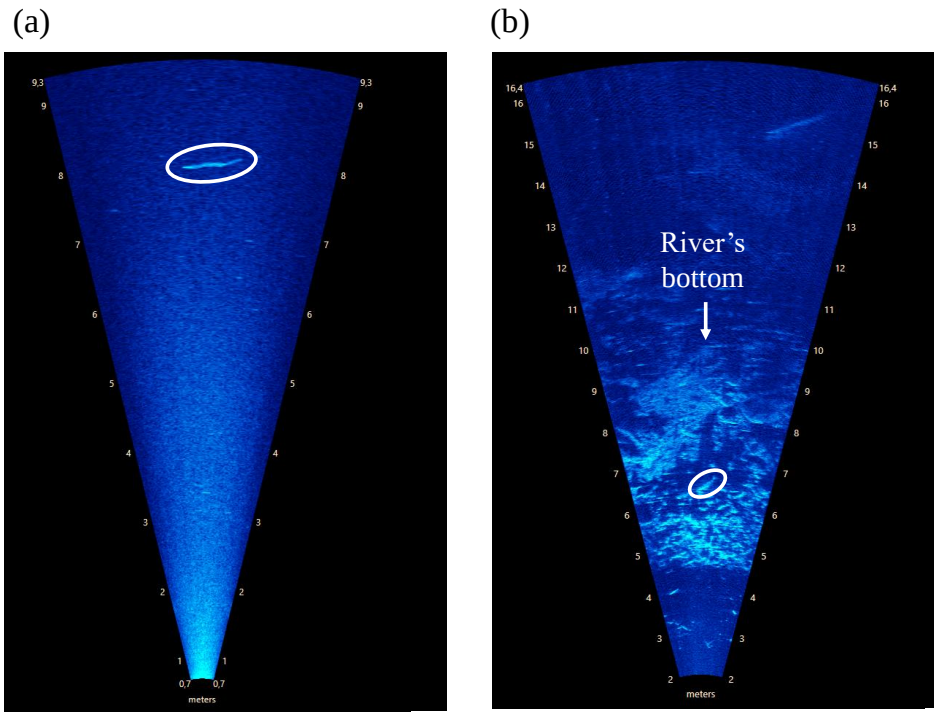


Figure 90 – Acoustic image recorded in open water at Mauzac using an ARIS 1800 with an eel's passage displayed (a), and a second image recorded at the Sélune monitoring site with the AC FOV hitting the river bottom with a passage of an individual from the CYP group of species.

All these factors that influence the automatic analysis of acoustic videos can also have an effect on the discrimination made by the operator during the viewing phase. This can have an impact on the operator's truth. In order to limit these potential biases as much as possible, it would be necessary to quantify the uncertainty of the operator's truth. This could be done by comparing the labelling carried out by two operators on the same set of videos, or by subjecting the same set of videos to the same operator several times.

2 Different methods to identify fish species

Two types of approaches have been used to identify species on acoustic videos. First, approaches based on a descriptor-classifier structure [Mueller et al., 2008, Bothmann et al., 2016], that extract target's features using image processing before classifying them using machine learning classifiers based on the average features measured on the whole trajectory in the AC FOV. Second, approaches based on deep learning methods classify the target into species classes at each detection before providing a final identification based on the most frequent class along the target's passage. Despite promising studies on acoustic videos analysis through deep learning methods [Zang et al., 2021, Kandimalla et al., 2022, Fernandez Garcia et al., 2023],

these approaches require a large amount of labelled data to cover the large variability of acoustic pictures a single species can produce. Moreover, they largely depend on morphological characteristics, AC settings and site configurations. Furthermore, the resolution of the images recorded by AC is poorer than the optical video resolution and may therefore decreases the capacity of a deep learning image-based analysis to identify species, by recording less detailed characteristics of the fish body. Hence, on a given acoustic image, only the overall morphology of the fish can be observed, without, or rarely with, details of the fins, making it almost impossible for an operator to make an identification from one single image. The operator needs to collect information from the whole target passage to provide an identification.

Before starting my PhD, we tried to use a machine learning classification model for the specific case of European eel identification from the AC data ([Le Quinio et al., 2022], Appendix B). This model was based on the Histogram of Oriented Gradients (HOG) descriptor paired with the Support Vector Machine (SVM) classifier. HOG descriptor focuses on the shape of an object [Dalal and Triggs, 2005]. SVM are linear classifiers based on the principle of the margin maximisation [Adankon and Cheriet, 2009]. However, the poor results we obtained have supported my decision to propose an advanced method of fish characteristics extraction from successive acoustic images, before using a machine learning classifier to carry out the classification process. Using a descriptor-classifier approach allows us to extract all features on which the human eye relies on to identify species: morphological characteristics, motion characteristics and the fish body deformation during its swimming behaviour. All these characteristics are calculated from the whole target's passage.

3 The proposed pipeline: the first step to an automatic and generic tool

In this PhD work, I got interested in developing methods, using an image-processing tool, (a) to automatically extract those features on acoustic videos and (b) to evaluate how they can promote fish species identification (Figure 86).

Three types of features can be extracted from acoustic videos: morphology, motion and swimming locomotion. I have considered these features in a complementary approach for multi-species identification. Hence, to correctly analyse the deformation of the target's body, the analysis of its motion must be done simultaneously to collect information on its velocity, trajectory and behaviour. Likewise, pairing the deformation of the body and the fish morphology would strengthen the identification process. Hence, we observe sometimes similar deformation patterns between eels and catfish, adding quantifiable information about their morphology and in particular their shape could help to distinguish those two species.

Additional environmental variables might also be used in the identification process. Hence, the water flow, turbidity, season and temperature are factors that are known to influence fish behaviour [Fenkes et al., 2016, Dahlke et al., 2020]. Further ecological knowledge of the species would be another source of information to consider. In the case of diadromous fish species, knowledge about migration periods or phenology would guide the identification process to distinguish species whose morphologies could be close, such as European eels and sea lampreys. Biological knowledge, such as feeding behaviour, direction or swimming speed may also improve species identification.

3.1 Extraction and interpretation of the morphological and motion features

In the literature, the features mostly extracted from acoustic videos are related to fish morphology and motion [Mueller et al., 2008, Bothmann et al., 2016]. Accordingly, the pipeline I first developed focuses on these two features. Both features allow a good distinction between fish and other objects, such as debris [Mueller et al., 2008, Bothmann et al., 2016]. Moreover, for species with a very characteristic shape such as eels, the method appears particularly suitable. Achieving an accurate description of the target's shape from the acoustic imaging, despite the reduced quality of the images, is a key step.

The proposed morphology and motion based method was tested on acoustic videos recorded at two different monitoring sites with eel populations of different sizes and with two types of AC. The objective was to determine if a single method could be applied to videos with different acquisition settings and distinct populations. The method achieved encouraging results on large European eels from ARIS data (recall = 74%, precision = 84%, N = 753), compared to previous studies [Mueller et al., 2008, Bothmann et al., 2016]. For the same species, [Bothmann et al., 2016] reached higher performance on DIDSON data (recall = 91%, precision = 96%) but their window length was twice shorter than the Mauzac one (5 meters vs. 10 meters) and the video resolution is higher without issue of target's image fragmentation. The performances obtained with our proposed method is slightly lower than the results of [Zang et al., 2021] on ARIS data (recall = 84%) using deep learning approaches for the identification of four individuals of American eels (vs 788 on our datasets), but there was no other fish species in their dataset, removing the risk of species confusion.

Both [Mueller et al., 2008] and [Bothmann et al., 2016] tested their method on data from a single type of AC and recorded in only one monitoring site. Our method performance is influenced by fish sizes, reaching poorer results on small European eels, mainly smaller than 60 cm, from the ARIS videos of the Port-La-Nouvelle monitoring site (recall = 44%, precision = 42%,

N = 788), highlighting that the number of pixels representing the target influences the ability to correctly identify it. However, our method performed identically on ARIS and BlueView videos for large eels (recall = 70%, precision = 74%, N = 198). Those results highlight the transferability of our method on two different types of AC, which has never been reported in the literature so far.

However, if a species such as eel has a very characteristic body shape on acoustic images, this is not the case for species such as Atlantic salmon (Figure 91.a) and pikeperch or Northern pike (Figure 91.b). In this case the benefit of the morphological analysis for species identification might be limited. Although providing valuable information, morphological and motion analysis therefore do not seem sufficient on their own to carry out a multi-species identification.

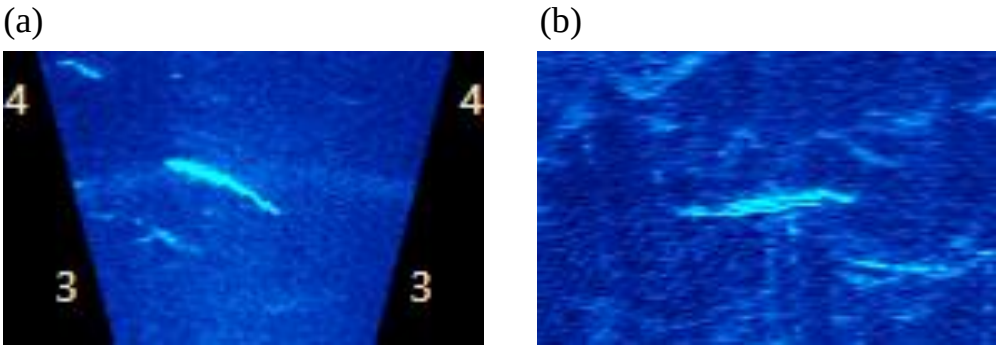


Figure 91 – Acoustic images of an Atlantic salmon (a) and of an individual from the PRED group of species (b).

3.2 Extraction of fish swimming mode from AC data: an innovative analysis

When an operator reviews acoustic videos, the swimming locomotion, i.e., how a fish deforms its body to swim, is one of the main pieces of information that allows him or her to distinguish a species from another. To our knowledge, exploiting the swimming locomotion of fish from acoustic videos has never been attempted yet, despite the relevant information it may provide for species identification.

Analysing the swimming locomotion of a target corresponds to monitor and study the deformation of the echo shapes from one frame to another. Image deformation is a vast field of computer vision widely used for 2D but also 3D image animation [Jain et al., 1998, Frisken and Mirtich, 1998]. The aim is to apply constraints to an image in order to deform its shape while keeping a coherent aspect to allow the eye to recognize it, these are the role of the deformation models.

As for the morphological analysis, the reliability of the extracted target's image plays a major role in the trustworthiness of the deformation calculation. If the binary image of the target is incorrect, i.e., if it aggregates pixels that do not correspond to the image of the target, the model will still tend towards the wrong pixels and the calculated deformation will be misrepresented. However, if the input data is correct, the generated map of deformation displays characteristic patterns making its analysis promising even if the main challenge remains to quantify this deformation.

To explore the discrimination ability of this information, the deformation maps of several individuals of the five identified groups of species have been processed and compared to each other. The preliminary results obtained show a strong similarity for individuals of the SIL group but a strong intra-group variability for the EEL and SAT groups. Many parameters can affect the fish swimming behaviour and should therefore be taken into account to correctly exploit these maps of deformation in the context of a multi-species identification. The environmental conditions such as water flow [Kirk et al., 2015, Lenihan et al., 2020], the behaviour of the fish (feeding, hunting) and its direction of migration (upstream, downstream) are all factors that influence the way the fish will deform its body to move.

Although the implementation of the deformation calculation method is complete, the exploitation of these deformation data still requires additional developments and additional tests on a larger dataset to reach more robust conclusions.

3.3 Additional features to strengthen the identification process

Exploiting data provided by the acoustic cameras cannot be done without taking into account the environmental conditions and in particular water velocity. Hence, the fish speed, orientation and behaviour highly rely on the water flow. At high water flow, a downstream fish will swim faster to cross the AC FOV, and its velocity and trajectory could be altered compared with low water flow. Similarly, the velocity and therefore the water flow influence how the fish deforms its body to swim. Moreover, if the fish swims in the direction of the current (downstream movement), the fish can drift, particularly in the case of high flows, or can actively swim. On the contrary, if the fish swims against the current (upstream movement), this swimming behaviour will differ and will depend on the water flow encountered.

Feeding, breeding or hunting behaviours of the fish are other factors influencing its swimming and therefore its motion and the way it deforms its body to swim. They are sources of intra-individual variability of the extracted features. Adding ecological knowledge to the classification process might therefore be relevant to correctly analyse them. Besides, for some

species, ecological knowledge will be the only information that allows discrimination. For example, both diadromous species, the European eel and the sea lamprey, have a similar shape and swimming mode making it difficult to distinguish them on acoustic videos. However, both species have different life cycles, they do not make their downstream and upstream migrations at the same period, the size of the downstream eels will differ from the ones of upstream lampreys and vice versa, allowing to classify their passages of the acoustic videos.

Beyond environmental and ecological knowledge, other information could be added to strengthen fish species identification. It may be interesting to pair the acoustic cameras with other types of sensors, recording complementary data. Within the frame of this PhD work, an experimental monitoring system was set up pairing two BlueView cameras with four split-beam sounders. The six devices equipped a navigation lock, with one AC and two split-beam sounders installed on each bank, perpendicularly to the current and uniformly placed to cover almost the whole water column. The objective was to detect fish on both AC and sounder data, providing simultaneously morphological and behavioural characteristics on AC data and the acoustic measure of its echoes, as well as the 3D position of the target on sounder data. The method I developed to retrieve information from the AC and the split-beam for the same target works well (Figure 92). However, because of operational issues and unexpected delays, the number of recorded passages of the species or groups of species studied in this PhD work was very low, less than ten, making it impossible to go further with this experiment. Nevertheless, this work has allowed me to gain experience in setting up and following up an experimental monitoring.

To carry out an automatic classification of species from all of the features collected, the development of a classification model is essential. It would help discard any empirically chosen parameter, as it is the case for the morphological and motion analysis proposed for anguilliform species. Hence, a classification model, in this case a supervised one (Figure 93), must be trained on a first set of data and evaluated on a second one to assess its ability to correctly label the data in classes. Once trained, the model is able to perform a class prediction on unseen data. To develop such a model, it is necessary to enlarge our datasets with more individuals, especially from the groups of species for which a low number of passages have been collected. These data should also ideally be recorded under various environmental conditions and with a good representation of their different behaviours.

Additional work is required to develop the classification model, that can be designed as described hereafter. At each detection of a target on the acoustic videos, all of the features are extracted and fill a matrix of size NxM with N the number of features, equal for each individual, and M the number of detections, which can vary from one individual to another. A first step would therefore be to standardise these matrices to a similar size by data reduction methods for example [Joliffe and Morgan, 1992, Izenman, 2008, Wang and Zhang, 2013]. These resulting matrices could then be used as input to a classification algorithm. Many machine learning

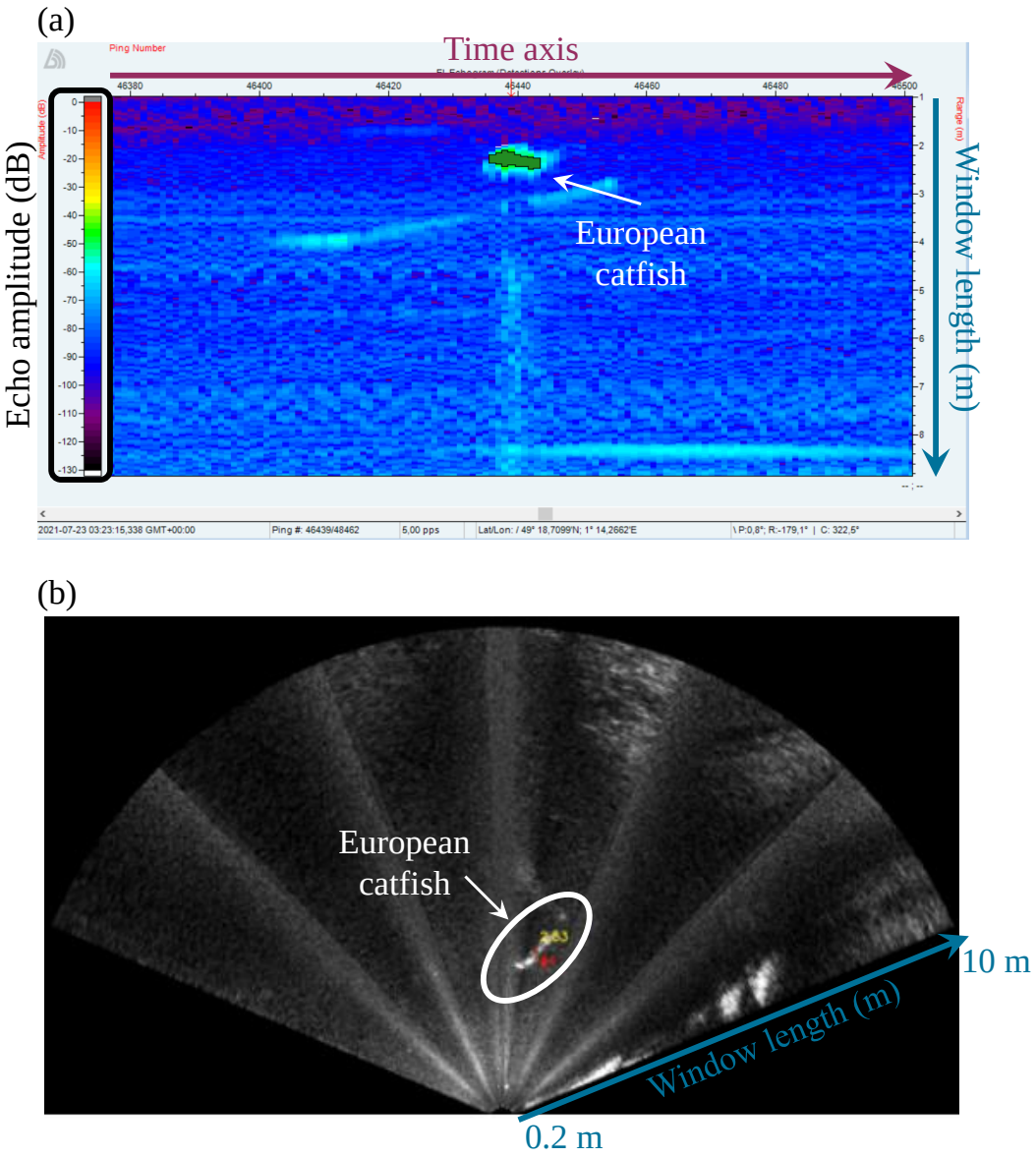


Figure 92 – Split-beam echogram (a) and acoustic image (b) of the corresponding European catfish.

algorithms are available to perform this classification step [Ray, 2019]. In order to integrate ecological knowledge into the classification, integrating *a priori* knowledge in the classification process should be discussed [Lauer and Bloch, 2008, Abbate et al., 2010].

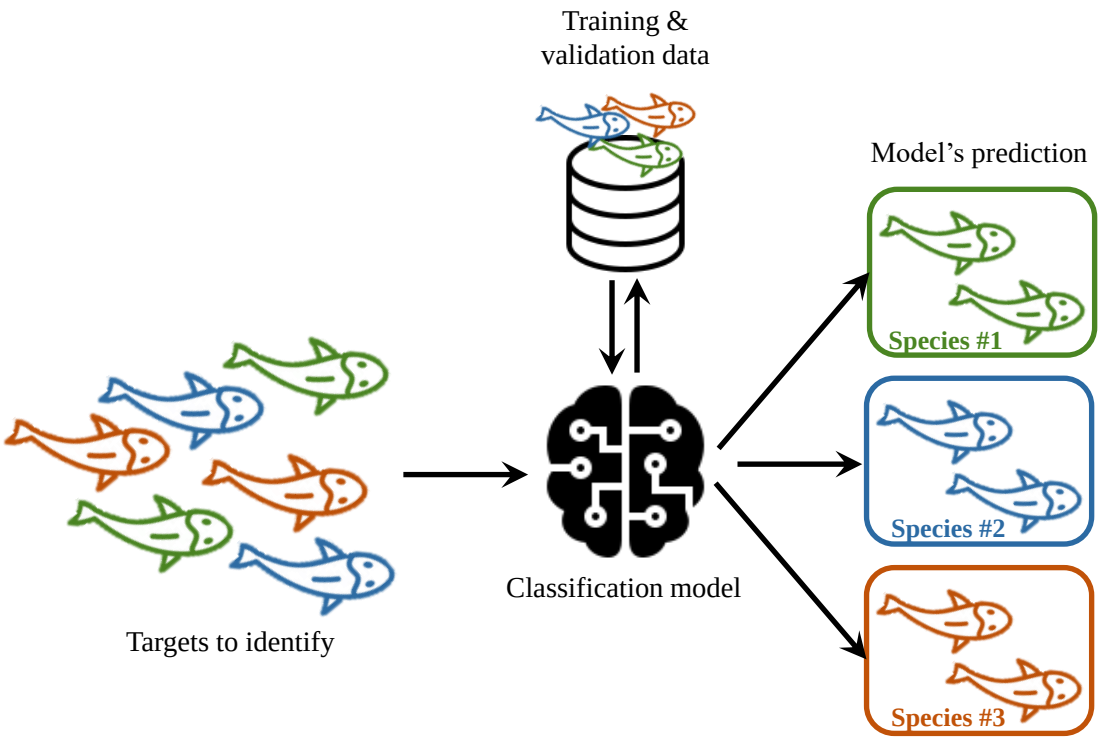


Figure 93 – Principle of a classification model.

4 Encouraging perspectives for fish population management

4.1 Acoustic videos at the service of mitigation measures optimization

Thanks to the proposed species identification method, the analysis of acoustic videos recorded over long periods of time to monitor fish populations can be improved. The case study of the Mauzac monitoring site highlights the valuable information AC can retrieve for facilities management. Hence, the morphology and motion based method allows to monitor the peaks of migration of the European eel over the three migration periods recorded by the AC at Mauzac. The information collected on the eels passages also give accurate information on their daily distribution and can lead to other complementary analyses to study the main factors governing the downstream migration peaks. From a management perspective, such information is useful for the development and implementation of the mitigation measures.

However, despite their ability to monitor for long periods of time without any manipulation of the fish, AC are limited by the range of their FOV. Hence, the study case of Mauzac has highlighted the importance of the AC position at the monitoring site and the importance of correctly exploiting the AC data, particularly by taking into account the configuration of the

monitoring site. Regarding Mauzac, when water flow increases, the eel counting on the acoustic videos largely decreases. The radio telemetry study [ECOGEA, 2014] highlights that in this high flow configuration, the eels favour a different passage that is no longer covered by the AC. No counting on the acoustic videos therefore does not mean that there are no eels passages at the monitoring site.

The automation of acoustic videos analysis in near-real time makes it possible to equip a monitoring site with several AC. The first limitation to such installation is that the AC should not interfere with each other while the second limitation is the purchasing cost of the AC. Multiple AC would cover more passage points and therefore obtain more information and a better proxy of fish passages counting. Another way would be to greatly increase the AC window length, by using certain types of AC at low-frequency mode. It will enable us to cover a distance up to 40 m for the ARIS Explorer 1800 and up to 100 m for the BlueView M900-2250. However, the target's identification at large window length would be harder because of the poor image resolution but detecting targets would be possible. By pairing those AC with the ones of shorter window length and by assuming that if fish are identified by the AC with the smaller window length, then targets seen at greater ranges are also considered as fish, it would then be possible to get a better proxy of the fish passages [Pratt et al., 2021].

Furthermore, the encapsulation of the morphology and motion method in a human-machine interface allows all managers to access the analysis of their acoustic videos. As of now this interface is only applicable to anguilliform species, our long-term goal is to integrate the entire proposed pipeline and complete it with the classification model in order to provide multi-species identification.

4.2 Collection of unrivalled information on fish populations

In addition to providing interesting information on the passage of fish populations, the proposed method collects information specific to each individual detected in the AC FOV. Fish length is one of the main information retrieved when visualising the acoustic videos. Its measure is known to be close to the real length of the fish despite potential operator bias [Burwen et al., 2010, Zhang et al., 2014, Lin et al., 2016, Daroux et al., 2019, Helminen et al., 2020]. In the future, the proposed morphological analysis can provide an automatic calculation of this fish length measured on the successive detections of the fish helping to reduce potential errors due to the fish position and orientation to the AC. In addition to providing an important information for species recognition [Wei et al., 2022], the method provides a fish length value that is free of operator bias. In order to have a correct idea of the accuracy of the length provided by the method, testing it on passages of individuals of known size should be done [Daroux et al., 2019, Lagarde et al., 2020].

The collection of the whole fish trajectory in the AC FOV also allows to get additional information on the fish swimming velocity and on its swimming behaviour: the fish is passing straight, staying around the same location or suddenly accelerating. The monitoring and quantification of the fish body deformation will also give additional information on the fish behaviour and can be seen as another approach to the fish tail beat analysis of [Mueller et al., 2010] although providing information on the whole body. Compiling all of these features will allow to categorise the fish behaviour: the fish is feeding, breeding, hunting or just migrating straight.

Thanks to AC properties and long-term monitoring, all of that biological information, that is hardly accessible with other methods, can be collected over long periods of time and respond to numerous and diverse ecological questions.

Furthermore, non-intrusive methods of fish natural behaviour description are rare, making therefore the automatic analysis of acoustic videos a promising tool and the information it provides very valuable.

5 Final words

My PhD work proposes a pipeline to extract all of the features available from acoustic videos: morphological, motion and swimming locomotion characteristics. If the extraction of morphological and motion features has already been exploited in the literature, the method proposed here works on two different types of AC, to detect and identify large anguilliform individuals, for now. However, the tools I have developed should also work on similar types of images such as those recorded by the DIDSON and Occulus cameras, the performance on that data must nevertheless be evaluated. The method on swimming behaviour automatic description is unique and innovative considering the existing literature about acoustic videos analysis. Although it is essential to deploy the pipeline on a much larger database, the prospects are very promising. Hence, it allows us to characterise the swimming mode of fish and therefore to study their behaviour. A further study in a controlled environment would allow a better understanding of the impact of the water flow on these swims and therefore a better reading of their body deformation when recorded in a natural environment while providing a reference for automatic species identification. Increasing the variability of the fish passages by multiplying the datasets from different monitoring sites would also help to maximise the performance of the automatic species identification.

To conclude, the pipeline outputs provide quantitative and qualitative information from acoustic videos that are necessary for aquatic environment managers to ensure the preservation of

fish populations. This work offers numerous perspectives for better understanding of the behaviour of the species studied, while offering the advantage of being a non-intrusive method. The pipeline I developed constitutes a milestone in the objective to have a fully automatic and real time tool giving long-term information on the fish populations present on a monitoring site. The design and implementation of the classification model will be the further step to reach this objective.

References

[Aalto et al., 2015] Aalto, E., Capoccioni, F., Terradez Mas, J., Schiavina, M., Leone, C., De Leo, G., and Ciccotti, E. (2015). Quantifying 60 years of declining European eel (Anguilla anguilla L., 1758) fishery yields in Mediterranean coastal lagoons. *ICES Journal of Marine Science*, 73(1):101–110.

[Abbate et al., 2010] Abbate, D., Asmundis, R., and Guarracino, M. (2010). *Prior Knowledge in the Classification of Biomedical Data*, volume 77, pages 1–8.

[Able et al., 2014] Able, K., Grothues, T., Rackovan, J., and Buderman, F. (2014). Application of mobile dual-frequency identification sonar (didson) to fish in estuarine habitats. *Northeastern Naturalist*, 21:192–209.

[Adankon and Cheriet, 2009] Adankon, M. and Cheriet, M. (2009). *Support Vector Machine*, pages 1303–1308.

[Ahmad and Sundararajan, 1987] Ahmad, M. O. and Sundararajan, D. (1987). A fast algorithm for two dimensional median filtering. *Circuits and Systems, IEEE Transactions on*, 34:1364 – 1374.

[Alex and Wahi, 2014] Alex, D. and Wahi, A. (2014). Bsfd: Background subtraction frame difference algorithm for moving object detection and extraction. *Journal of Theoretical and Applied Information Technology*, 60:623–628.

[Amundsen et al., 2003] Amundsen, P.-A., Dempson, J., Jonsson, B., Jonsson, N., O'Connell, M., and Mortensen, E. (2003). Atlantic salmon salmo salar l., brown trout salmo trutta l. and arctic charr salvelinus alpinus (l.): a review of aspects of their life histories. *Ecology of freshwater fish*, 12:1–59.

[April et al., 2021] April, J., Bardarson, H., Ahlbeck-Bergendahl, I., Bolstad, G., Breau, C., Buoro, M., Camara, K., Chaput, G., Cooper, A., Dauphin, G., Ensing, D., Erkinaro, J., Fiske, P., Freese, M., Gillson, J., Gregory, S., Hanson, N., Jepsen, N., Kelly, N., and Wennevik, V. (2021). Ices. 2021. working group on north atlantic salmon (wgnas).

[Artero et al., 2021] Artero, C., Marchetti, S., Bauer, E., Viala, C., Noel, C., Koenig, C., Berzins, R., and Lampert, L. (2021). High-resolution acoustic cameras provide direct and efficient assessments of large demersal fish populations in extremely turbid waters. *Applied Sciences*, 11:1899.

[Athanesious and Suresh, 2010] Athanesious, J. J. and Suresh, P. (2010). Systematic survey on object tracking methods in video.

[Baf et al., 2007] Baf, F., Bouwmans, T., and Vachon, B. (2007). Comparison of background subtraction methods for a multimedia learning space. pages 153–158.

[Baglinière and Porcher, 1994] Baglinière, J.-L. and Porcher, J. (1994). Caractéristiques des stocks de reproducteurs et comportement lors de la migration génésique. *Le Saumon Atlantique*, pages 101–122.

[Balaji and Karthikeyan, 2017] Balaji, S. R. and Karthikeyan, S. (2017). A survey on moving object tracking using image processing. In *2017 11th International Conference on Intelligent Systems and Control (ISCO)*, pages 469–474.

[Barbarossa et al., 2020] Barbarossa, V., Schmitt, R., Huijbregts, M., Zarfl, C., King, H., and Schipper, A. (2020). Impacts of current and future large dams on the geographic range connectivity of freshwater fish worldwide. *Proceedings of the National Academy of Sciences*, 117:201912776.

[Bay et al., 2006] Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. volume 3951, pages 404–417.

[Becker et al., 2013] Becker, A., Whitfield, A., Cowley, P., Järnegren, J., and Næsje, T. (2013). Potential effects of artificial light associated with anthropogenic infrastructure on the abundance and foraging behaviour of estuary-associated fishes. *Journal of Applied Ecology*, 50:43–50.

[Belcher et al., 2001] Belcher, E., Matsuyama, B., and Trimble, G. (2001). Object identification with acoustic lenses. volume 1, pages 6 – 11 vol.1.

[Bennett et al., 2020] Bennett, M., Becker, A., Gaston, T., and Taylor, M. (2020). Connectivity of large-bodied fish with a recovering estuarine tidal marsh, revealed using an imaging sonar. *Estuaries and Coasts*, 44.

[Bernatchez and Dodson, 1987] Bernatchez, L. and Dodson, J. (1987). Relationship between bioenergetics and behavior in anadromous fish migrations. *Canadian Journal of Fisheries and Aquatic Sciences*, 44:399–407.

[Blackman, 2004] Blackman, S. (2004). Multiple hypothesis tracking for multiple target tracking. *IEEE Aerospace and Electronic Systems Magazine*, 19(1):5–18.

[Bohlin et al., 1989] Bohlin, T., Hamrin, S., Heggberget, T., Rasmussen, G., and Saltveit, S. (1989). Electrofishing—theory and practice with special emphasis on salmonids. *Hydrobiologia*, 173:9–43.

[Boncelet, 2009] Boncelet, C. (2009). Chapter 7 - image noise models. In Bovik, A., editor, *The Essential Guide to Image Processing*, pages 143–167. Academic Press, Boston.

[Boswell et al., 2008] Boswell, K., Wilson, M., and Cowan, J. (2008). A semiautomated approach to estimating fish size, abundance, and behavior from dual-frequency identification sonar (didson) data. *North American Journal of Fisheries Management*, 28:799–807.

[Bothmann et al., 2016] Bothmann, L., Windmann, M., and Kauermann, G. (2016). Realtime classification of fish in underwater sonar videos. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65:n/a–n/a.

[Boulêtreau et al., 2018] Boulêtreau, S., Gaillagot, A., Carry, L., Tétard, S., De Oliveira, E., and Santoul, F. (2018). Adult atlantic salmon have a new freshwater predator. *PLOS ONE*, 13(4):1–12.

[Braga et al., 2022] Braga, L. T. M. D., Giraldo, A., and Godinho, A. L. (2022). Evaluation of three methods for manually counting fish in dam turbines using didson. *Hydrobiologia*, 849(2):309–321.

[Breder and Society, 1926] Breder, C. and Society, N. Y. Z. (1926). *The Locomotion of Fishes*. New York Aquarium nature series. New York Zoological Society.

[Brownscombe et al., 2019] Brownscombe, J., Lédée, E., Raby, G., Struthers, D., Gutowsky, L., Nguyen, V., Young, N., Stokesbury, M., Holbrook, C., Brenden, T., Vandergoot, C., Murchie, K., Whoriskey, K., Flemming, J., Kessel, S., Krueger, C., and Cooke, S. (2019). Conducting and interpreting fish telemetry studies: considerations for researchers and resource managers. *Reviews in Fish Biology and Fisheries*, 29.

[Brönmark et al., 2013] Brönmark, C., Hulthén, K., Nilsson, A., Skov, C., Hansson, L.-A., Brodersen, J., and Chapman, B. (2013). There and back again: Migration in freshwater fishes1. *Canadian Journal of Zoology*, 92:1–13.

[Burwen et al., 2010] Burwen, D., Fleischman, S., and Miller, J. (2010). Accuracy and precision of salmon length estimates taken from didson sonar images. *Transactions of the American Fisheries Society*, 139:1306–1314.

[Calonder et al., 2011] Calonder, M., Lepetit, V., and Fua, P. (2011). Brief: Binary robust independent elementary features.

[Capoccioni et al., 2019] Capoccioni, F., Leone, C., Pulcini, D., Cecchetti, M., Rossi, A., and Ciccotti, E. (2019). Fish movements and schooling behavior across the tidal channel in a mediterranean coastal lagoon: An automated approach using acoustic imaging. *Fisheries Research*, 219.

[Cappo et al., 2006] Cappo, M., Harvey, E., and Shortis, M. (2006). Counting and measuring fish with baited video techniques-an overview. *AFSB Conference and workshop "Cutting-edge technologies in fish and fisheries science"*, 1.

[Chapman et al., 2012] Chapman, B., Hulthén, K., Brodersen, J., Nilsson, A., Skov, C., Hansson, L.-A., and Brönmark, C. (2012). Partial migration in fishes: Causes and consequences. *Journal of fish biology*, 81:456–78.

[Cheong et al., 2019] Cheong, K. H., Poeschmann, S., Lai, J., Koh, J., Acharya, U. R., Yu, S., and Tang, K. (2019). Practical automated video analytics for crowd monitoring and counting. *IEEE Access*, 7:183252–183261.

[Cheung and Kamath, 2004] Cheung, S.-c. and Kamath, C. (2004). Robust techniques for background subtraction in urban traffic video. *Electronic Imaging International Society for Optics and Photonics*, pages 881–892.

[Cholley et al., 2015] Cholley, F., Perrin, T., and Vindimian, E. (2015). Expertise du projet d'effacement des ouvrages de Vezins et La Roche-qui-Boit sur la Sélune. Technical report, Ministère de l'écologie, du dévelopement durable et de l'énergie, Ministère de l'économie, de l'industrie et du numérique.

[Clay et al., 1998] Clay, P. M., Cowx, I., Evans, D., Gayanilo, F., Grainger, R., Gumy, A., Hongskul, V., Jarrett, T., Medley, P., Miyake, P., Pascoe, S., Riise, C., Sparre, P., Stamatopoulos, C., Venema, S., Vinther, M., Wan, T., and Zwieten, P. A. (1998). Guidelines for the routine collection of capture fishery data.

[Comaniciu et al., 2003] Comaniciu, D., Ramesh, V., and Meer, P. (2003). Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):564–577.

[Connolly et al., 2022] Connolly, R., Jinks, K., Shand, A., Taylor, M., Gaston, T., Becker, A., and Jinks, E. (2022). Out of the shadows: automatic fish detection from acoustic cameras. *Aquatic Ecology*.

[Cover and Hart, 1967] Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27.

[Cox and Hingorani, 1996] Cox, I. and Hingorani, S. (1996). An efficient implementation of reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(2):138–150.

[Cucherousset et al., 2018] Cucherousset, J., Horky, P., Slavík, O., Ovidio, M., Arlinghaus, R., Boulêtreau, S., Britton, R., García-Berthou, E., and Santoul, F. (2018). Ecology, behaviour and management of the european catfish. *Reviews in Fish Biology and Fisheries*, 28:177–190.

[Dahlke et al., 2020] Dahlke, F. T., Wohlrab, S., Butzin, M., and Pörtner, H.-O. (2020). Thermal bottlenecks in the life cycle define climate vulnerability of fish. *Science*, 369(6499):65–70.

[Dalal and Triggs, 2005] Dalal, N. and Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. In Schmid, C., Soatto, S., and Tomasi, C., editors, *International Conference on Computer Vision & Pattern Recognition (CVPR '05)*, volume 1, pages 886–893, San Diego, United States. IEEE Computer Society.

[Dambreville et al., 2006] Dambreville, S., Rathi, Y., and Tannenbaum, A. (2006). Tracking deformable objects with unscented kalman filtering and geometric active contours. In *2006 American Control Conference*, pages 6 pp.–.

[Daroux et al., 2019] Daroux, A., Martignac, F., Nevoux, M., Baglinière, J.-L., Ombredane, D., and Guillard, J. (2019). Manual fish length measurement accuracy for adult river fish using an acoustic camera (didson). *Journal of Fish Biology*, 95.

[De Oliveira and Le Quinio, 2022] De Oliveira, E. and Le Quinio, A. (2022). Suivi de la dynamique de dévalaison des anguilles argentées à l'aide d'une caméra acoustique - Développement d'une méthode de comptage automatique. EDF Internal report, Electricité de France (EDF).

[Dekker, 2016] Dekker, W. (2016). Management of the eel is slipping through our hands! distribute control and orchestrate national protection. *ICES Journal of Marine Science: Journal du Conseil*, 73:fsw094.

[Di Santo et al., 2021] Di Santo, V., Goerig, E., Wainwright, D., Akanyeti, O., Liao, J., Castro-Santos, T., and Lauder, G. (2021). Convergence of undulatory swimming kinematics across a diversity of fishes. *Proceedings of the National Academy of Sciences*, 118:e2113206118.

[Dingle, 1996] Dingle, H. (1996). *Migration: The Biology of Life on the Move*. Oxford University Press.

[Dingle and Drake, 2007] Dingle, H. and Drake, V. A. (2007). What Is Migration? *BioScience*, 57(2):113–121.

[dos Santos et al., 2017] dos Santos, M., Ribeiro, P. O., Núñez, P., Drews-Jr, P., and Botelho, S. (2017). Object classification in semi structured enviroment using forward-looking sonar. *Sensors*, 17:2235.

[Drouineau et al., 2018] Drouineau, H., Carter, C., Rambonilaza, T., Beaufaron, G., Bouleau, G., Gassiat, A., Lambert, P., Floch, S., Tétard, S., and De Oliveira, E. (2018). River continuity restoration and diadromous fishes: Much more than an ecological issue. *Environmental Management*, 61.

[Duda et al., 2001] Duda, R., Hart, P., and Stork, D. (2001). *Pattern classification*. Wiley, New York, 2nd edition.

[Dunkley and Shearer, 2006] Dunkley, D. and Shearer, W. (2006). An assessment of the performance of a resistivity fish counter. *Journal of Fish Biology*, 20:717 – 737.

[Durif et al., 2005] Durif, C., Dufour, S., and Elie, P. (2005). The silvering process of the eel: a new classification from the yellow resident stage to the silver migrating stage. *Journal of Fish Biology - J FISH BIOL*, 66:1025–1043.

[Durif and Elie, 2008] Durif, C. and Elie, P. (2008). Predicting downstream migration of silver eels in a large river catchment based on commercial fishery data fisheries management and ecology. *Fisheries Management and Ecology*, 15:127 – 137.

[ECOGEA, 2014] ECOGEA (2014). Evaluation de l'efficacité du dispositif de dévalaison au niveau de l'aménagement hydroélectrique EDF de Tuilières (Dordogne) pour les anguilles

d'avalaison. Résultats des tests d'automne-hiver 2012-2013. Technical Report E121015, ECOGEA.

[Eguiraun et al., 2014] Eguiraun, H., López-de Ipiña, K., and Martinez, I. (2014). Application of entropy and fractal dimension analyses to the pattern recognition of contaminated fish responses in aquaculture. *Entropy*, 16(11):6133–6151.

[Elgammal et al., 2002] Elgammal, A., Duraiswami, R., Harwood, D., and Davis, L. (2002). Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. *Proceedings of the IEEE*, 90(7):1151–1163.

[Engelbrecht and Berezovski, 2015] Engelbrecht, J. and Berezovski, A. (2015). Reflections on mathematical models of deformation waves in elastic microstructured solids. *Mathematics and Mechanics of Complex Systems*, 3:43–82.

[Fenkes et al., 2016] Fenkes, M., Shiels, H. A., Fitzpatrick, J. L., and Nudds, R. L. (2016). The potential impacts of migratory difficulty, including warmer waters and altered flow conditions, on the reproductive success of salmonid fishes. *Comparative Biochemistry and Physiology Part A: Molecular Integrative Physiology*, 193:11–21.

[Fernandez Garcia et al., 2023] Fernandez Garcia, G., Corpetti, T., Nevoux, M., Beaulaton, L., and Martignac, F. (2023). Acousticia, a deep neural network for multi-species fish detection using multiple models of acoustic cameras. *Aquatic Ecology*, pages 1–13.

[Floater, 2003] Floater, M. (2003). Mean value coordinates. *Computer Aided Geometric Design*, 20:19–27.

[French et al., 2018] French, G., Mackiewicz, M., Fisher, M., Challiss, M., Knight, P., Robinson, B., and Bloomfield, A. (2018). Jellymonitor: automated detection of jellyfish in sonar images using neural networks. pages 406–412.

[Friedman et al., 2021] Friedman, S., Price, S., and Wainwright, P. (2021). The effect of locomotion mode on body shape evolution in teleost fishes. *Integrative Organismal Biology*, 3.

[Frisken and Mirtich, 1998] Frisken, S. and Mirtich, B. (1998). A survey of deformable modeling in computer graphics.

[Gauthreaux, 1982] Gauthreaux, S. A. (1982). Chapter 2 – the ecology and evolution of avian migration systems.

[Gibson and Mirtich, 1997] Gibson, S. F. and Mirtich, B. (1997). A survey of deformable modeling in computer graphics.

[Golub and Reinsch, 1969] Golub, G. and Reinsch, C. (1969). Singular value decomposition and least squares solution. *Numerische Mathematik*, 14:42.

[Grote et al., 2014] Grote, A., Bailey, M., Zydlewski, J., and Hightower, J. (2014). Multibeam sonar (didson) assessment of american shad (alosa sapidissima) approaching a hydroelectric dam. *Canadian Journal of Fisheries and Aquatic Sciences*, 71:545–558.

[Guillard and Lebourges-Dhaussy, 2014] Guillard, J. and Lebourges-Dhaussy, A. (2014). À l'écoute des bancs de poissons. *Pour la science*, (436).

[Gurney et al., 2014] Gurney, W., Brennan, L., Bacon, P., Whelan, K., O'Grady, M., Dillane, E., and Mcginnity, P. (2014). Objectively assigning species and ages to salmonid length data from dual-frequency identification sonar. *Transactions of the American Fisheries Society*, 143:573–585.

[Han et al., 2009] Han, J., Honda, N., Asada, A., and Shibata, K. (2009). Automated acoustic method for counting and sizing farmed fish during transfer using didson. *Fisheries Science*, 75:1359–1367.

[Hayes et al., 2015] Hayes, J., Hay, J., Maxwell, I., and Quarterman, A. (2015). Estimating trout abundance with cataraft-mounted dual-frequency identification sonar: a comparison with drift diving. *North American Journal of Fisheries Management*, 35.

[Helminen et al., 2020] Helminen, J., Dauphin, G., and Linnansaari, T. (2020). Length measurement accuracy of adaptive resolution imaging sonar (aris) and a predictive model to assess adult atlantic salmon (salmo salar) into two size categories with long-range data in a river. *Journal of Fish Biology*, 97.

[Helminen and Linnansaari, 2021] Helminen, J. and Linnansaari, T. (2021). Object and behavior differentiation for improved automated counts of migrating river fish using imaging sonar data. *Fisheries Research*, 237:105883.

[Henkel et al., 2012] Henkel, C., Burgerhout, E., de Wijze, D., Dirks, R., Minegishi, Y., Jansen, H., Spaink, H., Dufour, S., Weltzien, F.-A., Tsukamoto, K., and Thillart, G. (2012). Primitive duplicate hox clusters in the european eel's genome. *PloS one*, 7:e32231.

[Herrero and Bescós, 2009] Herrero, S. and Bescós, J. (2009). Background subtraction techniques: Systematic evaluation and comparative analysis. In Blanc-Talon, J., Philips, W., Popescu, D., and Scheunders, P., editors, *Advanced Concepts for Intelligent Vision Systems*, pages 33–42, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Hightower et al., 2013] Hightower, J., Magowan, K., Brown, L., and Fox, D. (2013). Reliability of fish size estimates obtained from multibeam imaging sonar. *Journal of Fish and Wildlife Management*, 4:86–96.

[Hobson and Norris, 2008] Hobson, K. and Norris, R. (2008). Animal migration: A context for using new techniques and approaches. *Terrestrial Ecology*, 2:1–19.

[Hogan et al., 2014] Hogan, T., Cada, G., and Amaral, S. (2014). The status of environmentally enhanced hydropower turbines. *Fisheries*, 39.

[Holmes et al., 2006] Holmes, J., Cronkite, G., Enzenhofer, H., and Mulligan, T. (2006). Accuracy and precision of fish-count data from a "dual-frequency identification sonar" (didson) imaging system. *Ices Journal of Marine Science - ICES J MAR SCI*, 63:543–555.

[Huang et al., 2006] Huang, J., Shi, X., Liu, X., Zhou, K., Wei, L.-Y., Teng, S.-H., Bao, H., Guo, B., and Shum, H.-Y. (2006). Subspace gradient domain mesh deformation. *ACM Trans. Graph.*, 25:1126–1134.

[Huang et al., 2014] Huang, R., Han, J., and Tong, J. (2014). Assessment of fishery resource of a marine ranching based on a didson.

[Hue et al., 2002] Hue, C., Le Cadre, J.-P., and Perez, P. (2002). Tracking multiple objects with particle filtering. *IEEE Transactions on Aerospace and Electronic Systems*, 38(3):791–812.

[Hutchings and Jones, 1998] Hutchings, J. and Jones, M. (1998). Life history variation and growth rate thresholds for maturity in atlantic salmon, salmo salar. *Canadian Journal of Fisheries and Aquatic Sciences*, 55:22–47.

[Izenman, 2008] Izenman, A. J. (2008). *Linear Discriminant Analysis*, pages 237–280. Springer New York, New York, NY.

[Jacobs, 2005] Jacobs, D. (2005). Image Gradients. Class Notes for CMSC 426.

[Jain et al., 1998] Jain, A., Zhong, Y., and Dubuisson-Jolly, M.-P. (1998). Deformable template models: A review. *Signal Processing*, 71:109–129.

[Jebria et al., 2021] Jebria, N. B., Carmigniani, R., Drouineau, H., Oliveira, E. D., Tétard, S., and Capra, H. (2021). Coupling 3d hydraulic simulation and fish telemetry data to characterize the behaviour of migrating smolts approaching a bypass. *Journal of Ecohydraulics*, 0(0):1–14.

[Jing et al., 2016] Jing, D., Han, J., Wang, G., Wang, X., Wu, J., and Chen, G. (2016). Dense multiple-target tracking based on dual frequency identification sonar (didson) image. pages 1–5.

[Jing et al., 2017] Jing, D., Han, J., Wang, X., Wang, G., Tong, J., Shen, W., and Zhang, J. (2017). A method to estimate the abundance of fish based on dual-frequency identification sonar (didson) imaging. *Fisheries Science*, 83.

[Joliffe and Morgan, 1992] Joliffe, I. and Morgan, B. (1992). Principal component analysis and exploratory factor analysis. *Statistical methods in medical research*, 1:69–95.

[Jones et al., 2021] Jones, P. E., Champneys, T., Vevers, J., Börger, L., Svendsen, J. C., Consuegra, S., Jones, J., and Garcia de Leaniz, C. (2021). Selective effects of small barriers on river-resident fish. *Journal of Applied Ecology*, 58(7):1487–1498.

[Jonsson et al., 1990] Jonsson, B., Jonsson, N., and Hansen, L. (1990). Does juvenile experience affect migration and spawning of adult atlantic salmon? *Behavioral Ecology and Sociobiology*, 26:225–230.

[Justusson, 1981] Justusson, B. I. (1981). *Median Filtering: Statistical Properties*, pages 161–196. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Kandimalla et al., 2022] Kandimalla, V., Richard, M., Smith, F., Quirion, J., Torgo, L., and Whidden, C. (2022). Automated detection, classification and counting of fish in fish passages with deep learning. *Frontiers in Marine Science*, 8.

[Kang, 2011] Kang, M. (2011). Semiautomated analysis of data from an imaging sonar for fish counting, sizing, and tracking in a post-processing application. *Fisheries and aquatic sciences*, 14.

[Karami et al., 2017] Karami, E., Prasad, S., and Shehata, M. (2017). Image matching using sift, surf, brief and orb: Performance comparison for distorted images.

[Keefer et al., 2017] Keefer, M. L., Caudill, C. C., Johnson, E. L., Clabough, T. S., Boggs, C. T., Johnson, P. N., and Nagy, W. T. (2017). Inter-observer bias in fish classification and enumeration using dual-frequency identification sonar (didson): A pacific lamprey case study. *Northwest Science*, 91(1):41–53.

[Keeken et al., 2020] Keeken, O., Hal, R., Winter, H. V., Tulp, I., and Griffioen, A. B. (2020). Behavioural responses of eel (anguilla anguilla) approaching a large pumping station with trash rack using an acoustic camera (didson). *Fisheries Management and Ecology*, 27.

[Keith et al., 2021] Keith, P., Persat, H., Feunteun, E., and Allardi, J. (2021). *Les Poissons d'eau douce de France*. Collection Inventaires & Biodiversité. Biotope editions edition.

[Kelly and King, 2001] Kelly, F. and King, J. (2001). A review of the ecology and distribution of three lamprey species, lampetra fluviatilis (l.), lampetra planeri (bloch) and petromyzon marinus (l.): A context for conservation and biodiversity considerations in ireland. *Biology and Environment*, 101.

[Kennedy, 1985] Kennedy, J. (1985). *Migration: Behavioral and ecological*. ed.Migration:Mechanisms and Adaptive Significance. Contributions in Marine Science 27 (suppl.). Austin: Marine Science Institute, University of Texas.

[Kim and Bang, 2018] Kim, Y. and Bang, H. (2018). Introduction to kalman filter and its applications. In Govaers, F., editor, *Introduction and Implementations of the Kalman Filter*, chapter 2. IntechOpen, Rijeka.

[Kirk et al., 2015] Kirk, M. A., Caudill, C. C., Johnson, E. L., Keefer, M. L., and Clabough, T. S. (2015). Characterization of adult pacific lamprey swimming behavior in relation to environmental conditions within large-dam fishways. *Transactions of the American Fisheries Society*, 144(5):998–1012.

[Kulchandani and Dangarwala, 2015] Kulchandani, J. S. and Dangarwala, K. J. (2015). Moving object detection: Review of recent research trends. In *2015 International Conference on Pervasive Computing (ICPC)*, pages 1–5.

[Kupilik and Petersen, 2014] Kupilik, M. and Petersen, T. (2014). Acoustic tracking of migrating salmon. *The Journal of the Acoustical Society of America*, 136:1736–1743.

[Lagarde et al., 2021] Lagarde, R., Peyre, J., Amilhat, E., Bourrin, F., Prellwitz, F., Gael, S., and Elisabeth, F. (2021). Movements of non-migrant european eels in an urbanised channel linking a mediterranean lagoon to the sea. *Water*, 13:839.

[Lagarde et al., 2020] Lagarde, R., Peyre, J., Amilhat, E., Mercader, M., Prellwitz, F., Gael, S., and Elisabeth, F. (2020). In situ evaluation of european eel counts and length estimates accuracy from an acoustic camera (aris). *Knowledge and Management of Aquatic Ecosystems*, 421:44.

[Langkau et al., 2012] Langkau, M., Balk, H., Schmidt, M., and Borcherding, J. (2012). Can acoustic shadows identify fish species? a novel application of imaging sonar data. *Fisheries Management and Ecology*, 19:313–322.

[Larinier, 2001] Larinier, M. (2001). Environmental issues, dams and fish migration. *Dams, Fish and Fisheries: Opportunities, Challenges and Conflict Resolution.*

[Larinier, 2002] Larinier, M. (2002). Pool fishways, pre-barrages and natural bypass channels. *Bull. Fr. Pêche Piscic.*, 364:54–82.

[Larinier et al., 2005] Larinier, M., Matthieu, C., Frédérique, B., and Olivier, C. (2005). The use of radio telemetry for optimizing fish pass design. *Aquatic Telemetry: Advances and Applications.*

[Lauer and Bloch, 2008] Lauer, F. and Bloch, G. (2008). Incorporating prior knowledge in support vector machines for classification: A review. *Neurocomputing*, 71(7):1578–1594. Progress in Modeling, Theory, and Application of Computational Intelligenc.

[Le Quinio et al., 2023] Le Quinio, A., De Oliveira, E., Girard, A., Guillard, J., Roussel, J.-M., Zaoui, F., and Martignac, F. (2023). Automatic detection, identification and counting of anguilliform fish using in situ acoustic camera data: Development of a cross-camera morphological analysis approach. *PloS one*, 18:e0273588.

[Le Quinio et al., 2022] Le Quinio, A., Girard, A., Zaoui, F., and De Oliveira, E. (2022). Comparaison de deux méthodes d'identification et de comptage automatique des espèces anguilliformes à partir de caméras acoustiques. In *GRETSI*, Nancy.

[Lee and Park, 2012] Lee, J. and Park, M. (2012). An adaptive background subtraction method based on kernel density estimation. *Sensors*, 12(9):12279–12300.

[Lenihan et al., 2019] Lenihan, E., Mccarthy, T., and Lawton, C. (2019). Use of an acoustic camera to monitor seaward migrating silver-phase eels (anguilla anguilla) in a regulated river. *Ecohydrology Hydrobiology*, 19.

[Lenihan et al., 2020] Lenihan, E. S., McCarthy, T. K., and Lawton, C. (2020). Assessment of silver eel (anguilla anguilla) route selection at a water-regulating weir using an acoustic camera. *Marine and Freshwater Research*, 72(6):754–765.

[Lethlean, 1954] Lethlean, N. (1954). Xiii.—an investigation into the design and performance of electric fish-screens and an electric fish-counter. *Transactions of the Royal Society of Edinburgh*, 62.

[Li and Hong, 2014] Li, L. and Hong, J. (2014). Identification of fish species based on image processing and statistical analysis research. In *2014 IEEE International Conference on Mechatronics and Automation*, pages 1155–1160.

[Li, 2017] Li, S. (2017). A review of feature detection and match algorithms for localization and mapping. *IOP Conference Series: Materials Science and Engineering*, 231:012003.

[Li, 2019] Li, Z. (2019). Image metrics, a next-gen afm image analysis software.

[Liao, 2007] Liao, J. (2007). A review of fish swimming mechanics and behaviour in altered flows. *Phil. Trans. R. Soc. B*, 362:1973–1993.

[Limburg and Waldman, 2009] Limburg, K. E. and Waldman, J. R. (2009). Dramatic Declines in North Atlantic Diadromous Fishes. *BioScience*, 59(11):955–965.

[Lin et al., 2016] Lin, D.-Q., Zhang, H., Kang, M., and Wei, Q.-W. (2016). Measuring fish length and assessing behaviour in a high-biodiversity reach of the upper yangtze river using an acoustic camera and echo sounder. *Journal of Applied Ichthyology*, 32(6):1072–1079.

[Lindsey, 1978] Lindsey, C. (1978). *1 Form, Function, and Locomotory Habits in Fish*, volume 7, pages 1–100.

[Lowe, 2004] Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–.

[Lucas and Baras, 2000] Lucas, M. and Baras, E. (2000). Methods for studying spatial behaviour of freshwater fish in the natural environment. *Fish and Fisheries*, 1:283 – 316.

[Martignac, 2016] Martignac, F. (2016). *Utilisation de deux outils hydroacoustiques pour analyser la dynamique migratoire du saumon atlantique (Salmo salar L.) dans deux fleuves de la baie du Mont-Saint-Michel*. PhD thesis. Thèse de doctorat dirigée par Ombredane, Dominique Ecologie Rennes, Agrocampus Ouest 2016.

[Martignac et al., 2015] Martignac, F., Daroux, a., Baglinière, J.-L., Ombredane, D., and Guillard, J. (2015). The use of acoustic cameras in shallow waters: New hydroacoustic tools for monitoring migratory fish population. a review of didson technology. *Fish and Fisheries*.

[Mesejo et al., 2014] Mesejo, P., Valsecchi, A., Marrakchi-Kacem, L., Cagnoni, S., and Damas, S. (2014). Biomedical image segmentation using geometric deformable models and metaheuristics. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, 43.

[Mills, 1971] Mills, D. H. (1971). Salmon and trout: a resource,: Its ecology, conservation and management. Edinburgh, Scotland. Oliver and Boydn (Eds).

[Mizuno et al., 2015] Mizuno, K., Liu, X., Asada, A., Ashizawa, J., Fujimoto, Y., and Shimada, T. (2015). Application of a high-resolution acoustic video camera to fish classification: An experimental study. *2015 IEEE Underwater Technology, UT 2015*.

[Mora et al., 2018] Mora, E. A., Battleson, R. D., Lindley, S. T., Thomas, M. J., Bellmer, R., Zarri, L. J., and Klimley, A. P. (2018). Estimating the annual spawning run size and

population size of the southern distinct population segment of green sturgeon. *Transactions of the American Fisheries Society*, 147(1):195–203.

[Morais and Daverat, 2016] Morais, P. and Daverat, F. (2016). *An introduction to fish migration*. CRC Press.

[Mueller et al., 2010] Mueller, A.-M., Burwen, D. L., Boswell, K. M., and Mulligan, T. J. (2010). Tail-beat patterns in dual-frequency identification sonar echograms and their potential use for species identification and bioenergetics studies. *Transactions of The American Fisheries Society*, 139:900–910.

[Mueller et al., 2008] Mueller, A.-M., Mulligan, T., and Withler, P. (2008). Classifying sonar images: Can a computer-driven process identify eels? *North American Journal of Fisheries Management*, 28:1876–1886.

[ParksCanada, 2022] ParksCanada (2022). Restoring Atlantic salmon in the Clyburn Brook. https://www.pc.gc.ca/en/pnnp/ns/cbreton/decouvrirdiscover/conservation/saumon-salmon. Last checked on March 12, 2023.

[Parsons et al., 2017] Parsons, M., Fenny, E., Lucke, K., Osterrieder, S., Jenkins, G., Saunders, B., Jepp, P., and Parnum, I. (2017). Imaging marine fauna with a tritech gemini 720i sonar. *Acoustics Australia*, 45.

[Piccardi, 2004] Piccardi, M. (2004). Background subtraction techniques: a review. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, volume 4, pages 3099–3104 vol.4.

[Pornpanomchai et al., 2013] Pornpanomchai, C., Lurstwut, B., Leerasakultham, P., and Kitiyanan, W. (2013). Shape- and texture-based fish image recognition system. *Kasetsart Journal - Natural Science*, 47:624–634.

[Pratt et al., 2021] Pratt, T., Stanley, D., Schlueter, S., Rose, J., Weinstock, A., and Jacobson, P. (2021). Towards a downstream passage solution for out-migrating american eel (anguilla rostrata) on the st. lawrence river. *Aquaculture and Fisheries*, 6.

[Price et al., 2013] Price, V., Auster, P., and Kracker, L. (2013). Use of high-resolution didson sonar to quantify attributes of predation at ecologically relevant space and time scales. *Marine Technology Society Journal*, 47:33–46.

[Ramstad and Woody, 2003] Ramstad, K. M. and Woody, C. A. (2003). Radio tag retention and tag-related mortality among adult sockeye salmon. *North American Journal of Fisheries Management*, 23(3):978–982.

[Ray, 2019] Ray, S. (2019). A quick review of machine learning algorithms. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pages 35–39.

[Righton et al., 2016] Righton, D., Westerberg, H., Feunteun, E., Økland, F., Gargan, P., Amilhat, E., Metcalfe, J., Lobón-Cerviá, J., Sjöberg, N., Simon, J., Acou, A., Vedor, M., Walker,

A., Trancart, T., Brämick, U., and Aarestrup, K. (2016). Empirical observations of the spawning migration of european eels: The long and dangerous road to the sargasso sea. *Science Advances*, 2:e1501694.

[Rosenhahn et al., 2005] Rosenhahn, B., Kersting, U., Smith, A., Brox, T., Klette, R., and Seidel, H.-P. (2005). A silhouette based human motion tracking system. *http://citr.auckland.ac.nz/techreports/2005/CITR-TR-164.pdf*.

[Rourke et al., 2021] Rourke, M., Fowler, A., Hughes, J., Broadhurst, M., Dibattista, J., Fielder, D., Wilkes Walburn, J., and Furlan, E. (2021). Environmental dna (edna) as a tool for assessing fish biomass: A review of approaches and future considerations for resource surveys. *Environmental DNA*, 2022:9–33.

[Rublee et al., 2011] Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). Orb: an efficient alternative to sift or surf. pages 2564–2571.

[Rudstam et al., 2012] Rudstam, L. G., Jech, J. M., Parker-Stetter, S. L., Horne, J. K., Sullivan, P. J., and Mason, D. M. (2012). Fisheries Acoustics. In *Fisheries Techniques, Third edition*, page 40. A.v. zale, d.l. parrish, t.m. sutton edition.

[Saho, 2018] Saho, K. (2018). *Kalman Filter for Moving Object Tracking: Performance Analysis and Filter Design*.

[Sathyanarayanan et al., 2013] Sathyanarayanan, V., Mahadas, K., and Hung, G. (2013). A homeomorphic model of the effect of impact trauma on the human eye. *Journal of Computer Science and Systems Biololgy*, 6:317–326.

[Schaefer et al., 2006] Schaefer, S., McPhail, T., and Warren, J. (2006). Image deformation using moving least squares. *ACM Trans. Graph.*, 25:533–540.

[Selim and Koomullil, 2016] Selim, M. and Koomullil, R. (2016). Mesh deformation approaches – a survey. *Journal of Physical Mathematics*, 7.

[Senin, 2009] Senin, P. (2009). Dynamic time warping algorithm review.

[Sfakiotakis et al., 1999] Sfakiotakis, M., Lane, D., and Davies, J. (1999). Review of fish swimming modes for aquatic locomotion. *Oceanic Engineering, IEEE Journal of*, 24:237 – 252.

[Shahrestani et al., 2017] Shahrestani, S., Bi, H., Lyubchich, V., and Boswell, K. (2017). Detecting a nearshore fish parade using the adaptive resolution imaging sonar (aris): An automated procedure for data analysis. *Fisheries Research*, 191:190–199.

[Shen et al., 2023] Shen, W., Peng, Z., and Zhang, J. (2023). Identification and counting of fish targets using adaptive resolution imaging sonar. *Journal of fish biology*.

[Sibley et al., 2023] Sibley, E. C., Elsdon, T. S., Marnane, M. J., Madgett, A. S., Harvey, E. S., Cornulier, T., Driessen, D., and Fernandes, P. G. (2023). Sound sees more: A comparison of imaging sonars and optical cameras for estimating fish densities at artificial reefs. *Fisheries Research*, 264:106720.

[Silva et al., 2013] Silva, S., Servia, M., Vieira-Lanero, R., Barca-Bravo, S., and Cobo, F. (2013). Life cycle of the sea lamprey petromyzon marinus: Duration of and growth in the marine life stage. *Aquatic Biology*, 18:59–62.

[Simmonds and Maclennan, 2005] Simmonds, J. and Maclennan, D. (2005). Fisheries acoustics: Theory and practice: Second edition. *Fisheries Acoustics: Theory and Practice: Second Edition*, pages 1–252.

[Song et al., 2019] Song, C., Omalley, A., Roy, S., Barber, B., Zydlewski, J., and Mo, W. (2019). Managing dams for energy and fish tradeoffs: What does a win-win solution take? *Science of The Total Environment*, 669.

[Staines et al., 2022] Staines, G. J., Mueller, R. P., Seitz, A. C., Evans, M. D., O'Byrne, P. W., and Wosnik, M. (2022). Capabilities of an acoustic camera to inform fish collision risk with current energy converter turbines. *Journal of Marine Science and Engineering*, 10(4):483.

[Stauffer and Grimson, 1999] Stauffer, C. and Grimson, W. (1999). Adaptive background mixture models for real-time tracking. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, volume 2, pages 246–252 Vol. 2.

[Stein et al., 2015] Stein, F., Doering-Arjes, P., Fladung, E., Brämick, U., Bendal, B., and Schröder, B. (2015). Downstream migration of the european eel (anguilla anguilla) in the elbe river, germany: movement patterns and the potential impact of environmental factors. *River Research and Applications*, 32:666–676.

[Strachan et al., 1990] Strachan, N., Nesvadba, P., and Allen, A. (1990). Fish species recognition by shape-analysis of images. *Pattern Recognition*, 23:539–544.

[Syväranta et al., 2009] Syväranta, J., Cucherousset, J., Kopp, D., Martino L'Hostis, A., Cereghino, R., and Santoul, F. (2009). Contribution of anadromous fish to the diet of european catfish in a large river system. *Die Naturwissenschaften*, 96:631–5.

[Szabo-Meszaros et al., 2019] Szabo-Meszaros, M., Forseth, T., Baktoft, H., Fjeldstad, H.-P., T. Silva, A., Gjelland, K., Økland, F., Uglem, I., and Alfredsen, K. (2019). Modelling mitigation measures for smolt migration at dammed river sections. *Ecohydrology*, 12.

[Teichert et al., 2020] Teichert, N., Tétard, S., Trancart, T., Feunteun, E., Acou, A., and De Oliveira, E. (2020). Resolving the trade-off between silver eel escapement and hydropower generation with simple decision rules for turbine shutdown. *Journal of Environmental Management*, 261:110212.

[Terzopoulos, 1987] Terzopoulos, D. (1987). On matching deformable models to images. In *Topical Meeting on Machine Vision*, page FD1. Optica Publishing Group.

[Tesch, 2007] Tesch, F.-W. (2007). *The Eel: Third Edition*.

[Thibault, 1994] Thibault, M. (1994). Apercu historique sur l'evolution des captures et des stocks. In *Le saumon atlantique*. IFREMER.

[Thomsen and Willerslev, 2014] Thomsen, P. and Willerslev, E. (2014). Environmental dna – an emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, 183.

[Thorstad et al., 2013] Thorstad, E., Rikardsen, A., Alp, A., and Økland, F. (2013). The use of electronic tags in fish research–an overview of fish telemetry methods. *Turkish Journal of Fisheries and Aquatic Sciences*, 13:881–896.

[Thorstad et al., 2016] Thorstad, E., Todd, C., Uglem, I., Bjørn, P., Gargan, P., Vollset, K., Halttunen, E., Kålås, S., Berg, M., and Finstad, B. (2016). Marine life of the sea trout. *Marine Biology*, 163.

[Tiwari and Singhai, 2017] Tiwari, M. and Singhai, R. (2017). A review of detection and tracking of object from image and video sequences. *Int. J. Comput. Intell. Res*, 13(5):745–765.

[Travade and Larinier, 2002] Travade, F. and Larinier, M. (2002). Fish locks and fish lifts. *Bulletin Francais De La Peche Et De La Pisciculture*, 364.

[Tytell et al., 2010] Tytell, E., Borazjani, I., Sotiropoulos, F., Baker, T., Anderson, E., and Lauder, G. (2010). Disentangling the functional roles of morphology and motion in the swimming of fish. *Integrative and comparative biology*, 50:1140–54.

[Van der Walt et al., 2014] Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., and Yu, T. (2014). scikit-image: image processing in python. *PeerJ*, 2:e453.

[van Ginneken and Maes, 2005] van Ginneken, V. and Maes, G. (2005). The european eel (anguilla anguilla, linnaeus), its lifecycle, evolution and reproduction: A literature review. *Reviews in Fish Biology and Fisheries*, 15:367–398.

[Vautier et al., 2023] Vautier, M., Chardon, C., Goulon, C., Guillard, J., and Domaizon, I. (2023). A quantitative edna-based approach to monitor fish spawning in lakes: Application to european perch and whitefish. *Fisheries Research*, 264:106708.

[Viswanathan, 2011] Viswanathan, D. (2011). Features from accelerated segment test (fast).

[Vøllestad et al., 1986] Vøllestad, L., Jonsson, B., Hvidsten, N., Næsje, T., Haraldstad, , and Ruud-Hansen, J. (1986). Environmental factors regulating the seaward migration of european silver eels (anguilla anguilla). *Canadian Journal of Fisheries and Aquatic Sciences*, 43:1909–1916.

[Wang et al., 2014] Wang, S., Chu, L., Fu, Y., and Gao, W. (2014). An unfixed-elasticity mass spring model based simulation for soft tissue deformation. In *2014 IEEE International Conference on Mechatronics and Automation*, pages 309–314.

[Wang and Zhang, 2013] Wang, Y.-X. and Zhang, Y.-J. (2013). Nonnegative matrix factorization: A comprehensive review. *IEEE Transactions on Knowledge and Data Engineering*, 25(6):1336–1353.

[Webb et al., 2007] Webb, J., Verspoor, E., Aubin-Horth, N., Romakkaniemi, A., and Amiro, P. (2007). The atlantic salmon. *The Atlantic salmon: genetics, conservation and management*, 1:17–56.

[Webb, 1984a] Webb, P. (1984a). Body form, locomotion and foraging in aquatic vertebrates. *Integrative and Comparative Biology*, 24.

[Webb, 1984b] Webb, P. W. (1984b). Form and function in fish swimming. *Scientific American*, 251(1):72–83.

[Webb, 1988] Webb, P. W. (1988). Simple Physical Principles and Vertebrate Aquatic Locomotion1. *American Zoologist*, 28(2):709–725.

[Wei et al., 2022] Wei, Y., Duan, Y., and An, D. (2022). Monitoring fish using imaging sonar: Capacity, challenges and future perspective. *Fish and Fisheries*, 23.

[Welch and Bishop, 2006] Welch, G. and Bishop, G. (2006). An introduction to the kalman filter. *Proc. Siggraph Course*, 8.

[Weng et al., 2006] Weng, Y., Xu, W., Wu, Y., Zhou, K., and Guo, B. (2006). 2d shape deformation using nonlinear least squares optimization. *The Visual Computer*, 22:653–660.

[Wilson et al., 2014] Wilson, K., Allen, M., Ahrens, R., and Netherland, M. (2014). Use of underwater video to assess freshwater fish populations in dense submersed aquatic vegetation. *Marine and Freshwater Research*.

[Xu et al., 2009] Xu, C., Han, X., and Prince, J. L. (2009). Chapter 10 - gradient vector flow deformable models. In BANKMAN, I. N., editor, *Handbook of Medical Image Processing and Analysis (Second Edition)*, pages 181–194. Academic Press, Burlington, second edition edition.

[Xu et al., 2000] Xu, C., Pham, D., and Prince, J. (2000). Image segmentation using deformable models. *Handbook of Medical Imaging: Volume 2. Medical Image Processing and Analysis*.

[Xu et al., 2007] Xu, W., Zhou, K., Yu, Y., Tan, Q., Peng, Q., and Guo, B. (2007). Gradient domain editing of deforming mesh sequences. In *ACM SIGGRAPH 2007 Papers*, SIGGRAPH '07, page 84–es, New York, NY, USA. Association for Computing Machinery.

[Yin et al., 2020] Yin, T., Zang, X., Hou, Z., Jacobson, P., Mueller, R., and Deng, Z. (2020). Bridging the gap between laboratory and field experiments in american eel detection using transfer learning and convolutional neural network. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

[Yu et al., 2016] Yu, Z., Mizuno, K., Asada, A., Fujimoto, Y., and Shimada, T. (2016). New method of fish classification by using high-resolution acoustic video camera-aris and local invariant feature descriptor. pages 1–6.

[Zang et al., 2021] Zang, X., Yin, T., Hou, Z., Mueller, R., Deng, Z., and Jacobson, P. (2021). Deep learning for automated detection and identification of migrating american eel anguilla rostrata from imaging sonar data. *Remote Sensing*, 13:2671.

[Zhang et al., 2014] Zhang, H., Wei, Q., and Kang, M. (2014). Measurement of swimming pattern and body length of cultured chinese sturgeon by use of imaging sonar. *Aquaculture*, 434:184–187.

[Zhong et al., 2000] Zhong, Y., Jain, A., and Dubuisson-Jolly, M.-P. (2000). Object tracking using deformable templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(5):544–549.

[Zivkovic, 2004] Zivkovic, Z. (2004). Improved adaptive gaussian mixture model for background subtraction. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 2, pages 28–31 Vol.2.

[Zivkovic and Van der Heijden, 2006] Zivkovic, Z. and Van der Heijden, F. (2006). Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters*, 27:773–780.

Appendix A. Résumé de la thèse

1 Introduction

1.1 Migration et continuum écologique : un défi pour la préservation des populations de poissons

Les mouvements migratoires réalisés par les différentes populations de poissons leur permettent d'assurer des fonctions biologiques essentielles à leur cycle de vie [Dingle, 1996, Morais and Daverat, 2016]. La capacité de ces populations à migrer et à se déplacer est donc vitale et indispensable à leur survie. Cependant, les modifications d'origine anthropique des rivières, notamment à des fins de production énergétique, tendent à fragmenter les cours d'eau et à entraver les voies de migration et de déplacement des poissons. Pour rétablir la continuité écologique, des politiques de restauration et des mesures d'atténuation sont mises en œuvre (Directive-cadre sur l'eau 2000/60/CE pour la politique de l'eau du Parlement européen et du Conseil du 23 octobre 2000).

En tant que premier producteur d'hydroélectricité en France, EDF est un acteur important du continuum fluvial en France. L'entreprise doit ainsi mettre en œuvre des mesures d'atténuation afin de répondre aux exigences des politiques de restauration et préserver la biodiversité aquatique tout en optimisant les coûts de production.

Les migrations en amont et en aval doivent ainsi être assurées. Pour se faire, la suppression des sources de fragmentation est la solution la plus efficace mais elle n'est pas toujours possible, ni souhaitable. D'autres solutions de restauration, telles que la passe à poissons [Larinier, 2002], le canal de contournement naturel [Larinier, 2002], la capture et le transport, l'ascenseur à poissons [Travade and Larinier, 2002] assurent le franchissement des ouvrages par les poissons au cours de leur migration vers l'amont. En ce qui concerne la migration vers l'aval, différentes solutions ont été conçues, telles que la dérivation associée à des dispositifs de guidage (comme une crémaillère) ou la prise d'eau adaptée aux poissons. Cependant, en ce qui concerne les grandes centrales hydroélectriques, les seules solutions disponibles aujourd'hui sont des actions opérationnelles planifiées pour ouvrir le passage aux poissons. C'est le cas de la gestion des turbines pour les centrales hydroélectriques [Teichert et al., 2020, Song et al., 2019]. Des tur-

bines respectueuses des poissons peuvent également être installées [Hogan et al., 2014].

Afin d'assurer une bonne gestion de l'espace fluvial, d'évaluer l'efficacité des mesures d'atténuation ou en encore d'évaluer les stocks des espèces de poissons migrateurs, la mise en place de suivis des population de poissons est essentiel et permet d'apporter de précieuses informations aux différents acteurs impliqués tels qu'EDF.

1.2 L'imagerie acoustique : un outil non-intrusif pour le suivi des poissons

Parmi les nombreux outils disponibles pour mener à bien ces suivis piscicoles, les sonars multi-faisceaux à haute fréquence, appelés caméras acoustiques (CA), sont de plus en plus utilisées. Elles produisent des vidéos acoustiques permettant une observation directe de la morphologie du poisson, de sa tête à l'extrémité de sa caudale, ainsi que de ses mouvements et de son comportement. Ces outils permettent d'effectuer des suivis sur de longues périodes tout en étant peu dépendants des conditions environnementales [Martignac, 2016], notamment de la turbidité [Staines et al., 2022] et sont de plus totalement non intrusives. Plusieurs modèles de caméras acoustiques, aux propriétés différentes (résolution, volume couvert) sont disponibles.

Malgré ces nombreux avantages, des limites sont encore à surmonter pour exploiter pleinement les données des CA. Le stockage et le temps de traitement des vidéos acoustiques sont deux des principales contraintes inhérentes à ces outils. Le développement d'une méthode automatique et objective afin d'analyser l'imagerie acoustique à travers le comptage, l'identification et l'analyse du comportement des espèces est donc une étape clé à franchir afin d'extraire efficacement toutes les informations disponibles sur les vidéos.

1.3 L'analyse automatique de l'imagerie acoustique : une vue d'ensemble des méthodes existantes

Du comptage des poissons, à la mesure de leur taille et jusqu'à l'identification de leur espèce, chacune des méthodes d'analyse automatique proposées dans la littérature a été mise au point en fonction des besoins opérationnels des utilisateurs avec des approches et des applications qui leur sont propres. Parmi elles, l'identification automatique des espèces de poissons est l'un des principaux défis des vidéos acoustiques. Plusieurs études s'intéressent à distinguer des espèces à la forme très caractéristique comme les anguilles d'autres espèces ou de débris végétaux à l'aide de méthodes d'analyse morphologique [Mueller et al., 2008, Bothmann et al., 2016], ou de classification par apprentissage profond [Yin et al., 2020, Zang et al., 2021].

La mise au point de cette classification des espèces est difficile du fait de la nature des images

acoustiques. En effet, sur des images plus classiques de caméras optiques, un certain degré de détail est disponible pour identifier des espèces : couleurs, nageoires, aspect du corps et déformation du corps au cours de la nage. Sur les vidéos acoustiques beaucoup de ces détails ne sont pas disponibles. Seules la déformation du corps et la morphologie globale du poisson peuvent être visualisées, la faible résolution des images acoustiques ne permettant pas d'observer des critères morphologiques fins tels que le nombre ou le positionnement des nageoires.

1.4 Mode de nage et morphologie des poissons : des outils pour la classification des espèces

La déformation du corps du poisson se fait sous l'effet des forces physiques qui s'exercent sur lui dans un milieu aquatique [Lindsey, 1978]. La longueur d'onde et l'amplitude de l'onde propulsive sont les deux principales caractéristiques qui permettent de caractériser cette déformation à travers différents modes de nage [Sfakiotakis et al., 1999]. Quatre modes principaux de propulsion corporelle et/ou par nageoire caudale ont été identifiés par [Breder and Society, 1926] et [Lindsey, 1978] : les modes anguilliforme, subcarangiforme, carangiforme et thunniforme. Cette classification ne tient compte d'aucune affinité taxonomique, elle s'appuie uniquement sur le mode de nage des poissons.

Les espèces de poissons d'eau douce présentent également une grande diversité de morphologies, leur silhouette et leur longueur variant en fonction de l'espèce. La relation entre la forme globale des poissons et leur mode de nage a déjà été discutée dans la littérature et a souligné qu'elles sont toutes deux fortement corrélées [Webb, 1984a, Webb, 1984b, Webb, 1988, Tytell et al., 2010]. Outre la silhouette du corps, d'autres caractéristiques morphologiques sont utilisées pour l'identification des espèces, notamment la surface et la largeur [Strachan et al., 1990, Pornpanomchai et al., 2013, Li and Hong, 2014] des poissons. Cependant, l'un des principaux critères d'identification sur les vidéos acoustiques est la longueur de ces poissons [Martignac, 2016, Daroux et al., 2019].

Le mode de nage et la morphologie des poissons peuvent ainsi fournir des informations utiles pour l'identification des espèces. L'extraction automatique de ces informations à partir de vidéos acoustiques nécessite l'application de méthodes de vision par ordinateur initialement développées pour l'analyse de vidéos optiques et leur adaptation aux propriétés et à la résolution des vidéos acoustiques.

1.5 Objectifs de la thèse

A notre connaissance, aucune étude ne s'est intéressée à développer une méthode unique compatible avec des données enregistrées par différents modèles de CA sur différents sites de suivis ayant chacun leurs propres contraintes d'enregistrement et leur propre diversité d'espèces. L'objectif de cette thèse est ainsi de développer une solution opérationnelle compatible avec une analyse en temps réel et avec les données de différents modèles de caméras, afin de répondre à deux questions principales qui viseront à surmonter les principaux verrous technologiques de l'imagerie acoustique :

- Comment sélectionner et extraire uniquement les données d'intérêt parmi la grande quantité d'images acoustiques enregistrées, en détectant et suivant efficacement les poissons sur les vidéos ?
- Comment identifier les espèces de poissons à partir des informations disponibles sur les données d'imagerie acoustique ?

2 Matériels et méthodes

2.1 Imagerie acoustique

Les caméras acoustiques sont des sonars multifaisceaux à haute fréquence conçus pour créer des images à haute résolution [Martignac, 2016]. Les sonars multifaisceaux correspondent à une succession de n faisceaux uniques qui couvrent une section d'eau définie comme leur champ de vision. Lorsque l'impulsion acoustique rencontre une cible, un écho est renvoyé à la CA. La distance entre la caméra et la cible est calculée à partir du délai entre l'émission de l'impulsion acoustique et la réception de l'écho de la cible. L'ensemble de la morphologie de la cible ainsi que sa position dans le champ de vision sont donc visibles sur les images acoustiques grâce aux échos enregistrés dans chaque faisceau. Ainsi, les échos reçus par les n faisceaux adjacents cartographient le champ de vison de la CA, créant les images acoustiques en deux dimensions [Belcher et al., 2001].

Ce travail de thèse utilise les données enregistrées par trois de modèles de caméra acoustique : ARIS Explorer 1800 (Sound Metrics Corp., Bellevue, WA, USA), ARIS Explorer 3000 (Sound Metrics Corp., Bellevue, WA, USA) et BlueView M900-2250-130 2D (Teledyne Technologies Inc., Thousand Oaks, CA, USA). La principale différence entre les caméras ARIS et BlueView est l'ouverture de leur champ de vision (30° pour les caméras ARIS, 130° pour les caméras BlueView).

2.2 Les sites de suivi

Les données utilisées dans cette étude ont été enregistrées sur trois sites de suivi situés en France : le barrage de Mauzac, le site de suivi de Ducey et le site de suivi de Port-La-Nouvelle. Le barrage de Mauzac, situé sur la Dordogne, fait l'objet d'un suivi des anguilles européennes (*Anguilla anguilla*) lors de leur dévalaison. L'anguille européenne étant une espèce à fort intérêt écologique, figurant sur la liste rouge de l'UICN, des arrêts de turbinage sont effectués à Mauzac comme mesure d'atténuation. Le suivi de leurs passages à cet endroit est essentiel pour EDF, gestionnaire de l'ouvrage, afin de détecter les individus lors de leur migration vers la mer. Une caméra ARIS 1800 a enregistré en continu pendant trois périodes de migration tandis qu'une caméra BlueView a été installée en simultané pour la dernière période de migration.

Le site de suivi de Ducey, situé sur la Sélune, se trouve en aval des deux anciens barrages de La Roche-Qui-Boit et de Vezins. Ces deux barrages ont été récemment supprimés (2018-2022) afin de restaurer la continuité écologique [Cholley et al., 2015]. Les CA sont utilisés à Ducey pour analyser la dynamique migratoire des espèces diadromes avant et après la destruction des barrages [Martignac, 2016]. Dans ce cadre, une caméra ARIS 3000 a été installée pour quelques mois sur deux années différentes.

Les données enregistrées sur un troisième site de suivi ont été utilisées pour évaluer les performances de la méthode développée dans le cadre de cette thèse, étude publiée dans PLOS ONE [Le Quinio et al., 2023]. Ce site, secondaire, est situé à Port-La-Nouvelle et est équipé d'une caméra ARIS 1800, installée pour enregistrer les passages des anguilles européennes pendant leur migration vers l'aval. Les objectifs du suivi, les caractéristiques du site et les réglages de la CA sont détaillés dans [Lagarde et al., 2020] et [Lagarde et al., 2021].

2.3 Visualisation des données CA : la vérité de l'opérateur

Pour concevoir et évaluer une méthode d'identification des espèces, des données préalablement analysées par des opérateurs experts sont nécessaires. Dans le cadre de ce travail, aucune vérité terrain n'a pu être réalisée car aucune observation directe des poissons n'est possible, seul les enregistrements via les caméras acoustiques sont disponibles.

Cependant, il est possible d'extraire des informations morphologiques et comportementales des passages de poissons, ainsi qu'une identification spécifique, à partir des vidéos acoustiques grâce à leur visualisation par un opérateur. Cette analyse sera considérée comme la référence pour le travail proposé et est appelée la vérité opérateur. L'identification de l'espèce du poisson par l'opérateur repose sur des caractéristiques morphologiques et comportementales. Cependant, dans certains cas, ces caractéristiques sont communes à plusieurs espèces, ce qui empêche toute distinction entre elles. Ces espèces sont donc regroupées sous une classe commune.

Cinq groupes d'espèces sont considérés dans le cadre de ce travail de thèse : EEL, composé d'anguilles Européennes (*Anguilla anguilla*), SAT, composé de saumons atlantiques (*Salmo salar*), SIL, composé de silures glanes (*Silurus glanis*), CYP, composé de brèmes communes (*Abramis brama*) et de carpes communes (*Cyprinus carpio*), et enfin PRED, composé de brochets (*Esox lucius*) et de sandres (*Sander lucioperca*).

2.4 Jeux de données

L'analyse des vidéos enregistrées sur les sites de suivi par des opérateurs expérimentés définit trois jeux de données enregistrés sur les sites de suivi de Mauzac et de la Sélune, avec deux modèles de caméras acoustiques différents : MZC-ARIS (Mauzac, ARIS 1800), MZC-BV (Mauzac, BlueView) et SEL-ARIS (Sélune, ARIS 3000). Les jeux de données MZC-ARIS et MZC-BV sont spécifiques à l'anguille européenne car le suivi de Mauzac se concentre sur cette espèce, tandis que le jeu de données SEL-ARIS répertorie les passages de plusieurs espèces ou groupes d'espèces.

Ces jeux de données sont utilisés pour concevoir et évaluer la méthode d'analyse automatique des vidéos acoustiques proposée dans ce manuscrit.

3 Trier les données d'intérêt : détection et suivi, la première étape vers l'identification des espèces

Dans un premier temps, nous nous concentrons sur le tri des données d'intérêt à partir de la grande quantité de vidéos enregistrées par les CA. Dans le cas des études sur la migration piscicole, les données qui nous intéressent sont les cibles en mouvement pouvant potentiellement être des poissons. Une cible est définie comme une zone de pixels contigus d'intensités similaires qui peut être suivie d'une image à l'autre, le long de sa trajectoire dans le champ de vision de la caméra. Le processus de traitement, en trois étapes, est le suivant :

- **Extraction des séquences d'images d'intérêt (FOI).**
  La première étape consiste à distinguer les FOI dans lesquelles passent les objets d'intérêt (OOI). A cette étape du processus, les objets d'intérêt (OOI) correspondent à des régions de pixels contigus d'intensités similaires provenant du premier plan de l'image et se déplaçant d'une image à l'autre
- **Extraction des régions d'intérêt (ROI).**
  La deuxième étape de notre processus vise à filtrer les OOI présents dans les FOI. De plus, afin d'assurer un suivi efficace, les OOI sélectionnées sont post-traités pour être représentés sous la forme d'une seule région continue de pixels. Les images résultantes

correspondent aux ROI.

- **Extraction du passage des cibles.**
  Les ROI sont ensuite suivies d'une image à l'autre tout au long de leur trajectoire dans le champ de vision de la caméra. Chaque suivi complété devient une cible qui peut correspondre à un poisson.

3.1 Des vidéos acoustiques entières aux frames d'intérêt

Pour extraire les FOI d'une vidéo, il est nécessaire d'identifier les intervalles d'images pendant lesquels un OOI se déplace dans le champ de vision de la caméra. Chaque passage d'un OOI crée une discontinuité qui correspond à une intensité de pixel différente de celle des pixels d'arrière-plan. La méthode que nous avons développée synthétise l'ensemble de la vidéo sous la forme d'un échogramme dans lequel seules les discontinuités sont présentes. Pour évaluer la performance de cette étape, les valeurs de rappel, i.e., le nombre d'anguilles détectées par rapport au nombre total d'anguilles présentes, ainsi que les valeurs d'efficacité, i.e, le nombre de FOI par rapport au nombre total d'images de la vidéo, sont calculées.

Après un pré-traitement visant à débruiter les images, celles-ci sont analysées par balayage via des fenêtres coulissantes et chevauchantes à 50%. Deux approches ont été développées pour extraire l'information utile de ces fenêtres. La première se concentre sur les espèces anguilliformes pour lesquelles une décomposition en valeurs singulières est effectuée pour chacune des fenêtres glissantes. Cette méthode performe bien sur les vidéos ARIS (rappel = 92% ; efficacité = 83.1% ; N=25) et presque aussi bien sur les vidéos BlueView (rappel = 92% ; efficacité = 64.1% ; N = 25). Cependant, le temps de calcul est très élevé (jusqu'à 2.5 fois la durée de la vidéo dans le cas de la BlueView).

La seconde méthode développée a pour but de fournir une approche générique, non limitée aux espèces anguilliformes, et avec un temps de calcul plus faible pour permettre une analyse des vidéos en temps réel. Celle-ci repose sur le calcul de la surface des OOI. La méthode fournit des performances similaires à la première approche pour les vidéos ARIS (rappel = 100% ; efficacité 84.7% ; N = 25) et meilleure pour les vidéos BlueView (rappel = 100% ; efficacité = 76.9% ; N = 25). Testée sur le jeu de donnée multi-espèces, les résultats sont aussi encourageants (rappel = 90% ; efficacité = 86.1% ; N = 20). De plus, le temps de calcul est plus faible : il est inférieur ou égal à la durée de la vidéo étudiée.

3.2 Des frames d'intérêt aux régions d'intérêt

Les FOI étant extraits des vidéos acoustiques, l'étape suivante du processus vise à extraire automatiquement les images correspondantes pour y détecter les discontinuités, i.e., les ROI, à l'aide de méthodes de traitement d'image.

Un premier traitement est appliqué à l'image : un filtre médian, avec un élément structurant plus petit (fixé empiriquement à 50 mm), est appliqué pour lisser l'image avant d'appliquer le filtre de soustraction d'arrière-plan permettant d'extraire les objets en mouvement. L'étape de soustraction de l'arrière-plan conduit souvent à des images binaires fragmentées. Pour résoudre ce problème, nous effectuons une dilatation mathématique de l'image. Un élément structurant est choisi délibérément grand pour surmonter ce problème, correspondant à un rectangle de dimension 15x15 pixels. Un filtrage des ROI est ensuite effectué pour éliminer les plus petites zones qui ne peuvent pas correspondre aux poissons ciblés.

Afin de maximiser l'information de l'image pour les étapes suivantes du processus, nous proposons également une restauration des images ROI à travers une succession de filtres morphologiques mathématiques.

3.3 Des régions d'intérêt aux cibles

Nous souhaitons détecter les cibles à partir de toutes les ROI présentes dans une image à t=k, ce qui signifie que nous souhaitons suivre les ROI tout au long de leur trajectoire dans le champ de vision de la caméra. Les détections successives d'une même zone d'intérêt définissent une cible.

Deux méthodes de suivi ont été mises en œuvre. La première, très classique et simple d'implémentation, est la méthode de suivi de centroïde. Celle-ci n'est néanmoins pas adaptée dans le cas de nombreux passages simultanés. Un filtre de Kalman a donc été implémenté. Le filtre de Kalman est un algorithme qui estime les valeurs des variables mesurées dans le temps [Kim and Bang, 2018]. Cela signifie qu'il permet de prédire l'état d'un processus sur la base de ses précédentes valeurs. Ce filtre permet ainsi de corriger les erreurs de la première méthode bien que des tests supplémentaires sur un jeu de données plus large est nécessaire.

3.4 Conclusion

La méthode proposée peut être utilisée comme un moyen semi-automatique d'analyser les vidéos acoustiques : l'extraction automatique des cibles permet à un opérateur de ne seulement visualiser que ces séquences et non pas les chroniques entières afin de pouvoir ensuite identifier les espèces, mesurer les individus et effectuer un comptage des espèces cibles. La méthode peut également être une étape préliminaire à une analyse automatique des vidéos acoustiques. Elle vise alors à pré-traiter les données pour faciliter le processus d'identification de l'espèce.

4 Morphologie du corps et analyse du mouvement : des outils pour l'identification des espèces

4.1 Une approche d'analyse morphologique inter-caméras : l'étude de cas des espèces anguilliformes

Comme évoqué précédemment, les espèces anguilliformes ont une forme serpentine caractéristique qui en font un cas intéressant pour étudier l'identification des espèces sur des vidéos acoustiques à l'aide de l'analyse des caractéristiques morphologiques des poissons.

Si des méthodes automatiques aux résultats encourageants ont déjà été mises au point dans la littérature, celles-ci ont été testées sur des données enregistrées à faible distance (6 mètres, [Bothmann et al., 2016], ou sur un très faible nombre d'individus (N=5, [Zang et al., 2021]), toutes sur un seul modèle de caméra. Notre objectif est donc de mettre au point une méthode d'identification et de comptage des espèces anguilliformes transposable à différents sites de suivi et à partir des données enregistrées par différents modèles de caméras acoustiques.

Notre méthode repose dans un premier temps sur les différentes étapes de traitements permettant la détection des cibles d'intérêt, présentées précédemment. L'analyse de la morphologie et du mouvement est ensuite réalisée. Une nouvelle approche permettant de caractériser la silhouette de la cible est également proposée pour compléter ces informations. De plus, la taille des individus est automatiquement calculée et prise en compte dans le processus de classification.

La performance de la méthode a été évaluée sur les deux jeux de données spécifiques aux anguilles enregistrés à Mauzac (MZC-ARIS et MZC-BV) ainsi que sur le jeu de données supplémentaire de Port-La-Nouvelle à partir des valeurs de rappel et de précision. Des performances encourageantes sont atteintes grâce à cette méthode pour l'identification et le comptage des espèces anguilliformes à partir des vidéos acoustiques. De plus des résultats similaires sont obtenus sur les jeux de données de MZC-ARIS (rappel = 73.8%, précision = 83.5%, N = 753) et MZC-BV (rappel = 70.2%, précision = 74.3%, N = 198), ce qui met en évidence la capacité de la méthode à fonctionner sur des données enregistrées par différents type de modèles de CA.

Néanmoins, les performances atteintes sur le jeu de données de Port-La-Nouvelle sont largement inférieures (rappel = 44.3%, précision = 41.9%, N=788). La population d'anguilles européennes de Port-La-Nouvelle est principalement composée d'anguilles de plus petite tailles (40-60 cm) que celle passant à Mauzac (70-90 cm). Ces résultats soulignent la relation entre la performance de la méthode et la taille des individus. [Fernandez Garcia et al., 2023] met en évidence des conclusions similaires pour l'identification des poissons à l'aide de réseaux de neurones convolutifs (CNN). En effet, les cibles de petite taille seront définies par moins de pixels que les

cibles plus grandes, ce qui limite les informations morphologiques disponibles. Plus le nombre de pixels définissant la cible est élevé, plus l'analyse morphologique est fiable. Cette étude a fait l'objet d'une publication dans PLOS ONE [Le Quinio et al., 2023].

4.2 Cas d'étude : les anguilles européennes de la centrale hydroélectrique de Mauzac

La centrale hydroélectrique de Mauzac, exploitée par EDF, est le premier obstacle rencontré par les anguilles européennes du haut du bassin de la Dordogne lors de leur dévalaison. Ainsi, une obligation réglementaire impose l'arrêt des turbines pendant la période de dévalaison des anguilles. Ces arrêts sont planifiés à partir d'un modèle de prévision des migrations [E. De Oliveira, communication personnelle]. Afin de valider et optimiser ce modèle ainsi que de collecter des informations biologiques sur la migration des anguilles, le site de Mauzac est équipé, depuis 2014, de CA.

La méthode développée dans le cadre de cette thèse visant à identifier et compter les espèces anguilliformes a ainsi été appliquée sur la totalité des vidéos ARIS enregistrées à Mauzac. Les comptages obtenus permettent d'observer des pics de migration qui sont en majorité en concordance avec les prédictions du modèle utilisées pour planifier les arrêts de turbinage. Certains comptages, en faible nombre, ne coïncident cependant pas avec ces prédictions et sont certaines fois dus à des artefacts (phénomènes de arcing) qui sont en effet une des sources principales d'erreurs de notre méthode sur les vidéos ARIS. Ceux-ci déforment l'image de la cible à identifier en lui donnant une forme plus allongée ce qui vient biaiser sa classification. De plus, dans le cas de débits supérieurs au débits d'équipement de l'usine hydroélectrique, aucune anguille n'est comptée par la méthode. Ce résultat, conforme à nos attentes, s'explique par l'ouverture progressive des vannes du barrage, par lesquelles les anguilles passent préférentiellement dans ces conditions.

La mise au point de notre méthode de comptage automatique et en temps réel, permettant de s'affranchir des contraintes d'une visualisation des vidéos par un opérateur, rend possible l'équipement du site de Mauzac avec d'autres caméras acoustiques afin de couvrir davantage de points de passages et ainsi obtenir une meilleure estimation du nombre d'anguilles dévalant à Mauzac.

Afin de permettre l'utilisation de notre méthode par un plus grand nombre d'utilisateurs sans besoin de connaissances informatiques spécifiques, les algorithmes que j'ai développés ont été intégré dans une interface homme-machine. L'interface a été testée en laboratoire avec des vidéos enregistrées en temps réel et sera prochainement déployée dans un environnement opérationnel.

5 Le comportement de nage : un outil complémentaire pour l'identification des espèces

La déformation du corps d'un poisson au cours de son déplacement est une autre caractéristique d'intérêt qui peut être extraite des vidéos acoustiques afin de discriminer les espèces. En plus des caractéristiques morphologiques analysées grâce à la méthode présentée précédemment, celle-ci pourrait aider à améliorer la capacité d'identification des espèces. Pour extraire les informations correspondantes des vidéos, la déformation du corps du poisson doit être suivie et quantifiée à chaque pas de temps. Cela signifie que pour un même point du corps, nous souhaitons mesurer l'écart de position entre deux pas de temps successifs, indépendamment du déplacement du poisson. Pour ce faire, nous utilisons des modèles déformables. Les modèles déformables sont des courbes ou des surfaces définies à partir d'une image d'entrée. Ils peuvent se déplacer sous l'influence de forces internes provenant du modèle lui-même et de forces externes calculées à partir de l'image traitée [Xu et al., 2009]. En d'autres termes, cela signifie que les modèles déformables sont conçus pour être dirigés par des contraintes externes provenant de l'image traitée qui commandent la déformation, tout en préservant la forme initiale interne établie à partir de l'image d'entrée. La mise en œuvre d'un modèle déformable est la première étape pour décrire la déformation du corps du poisson. La seconde étape à réaliser est le traitement de cette déformation pour en extraire des caractéristiques utiles à la discrimination des espèces.

Nous ne cherchons pas à développer un modèle de classification sur la base de la déformation du corps du poisson. L'objectif est de développer ultérieurement un modèle de classification global avec toutes les informations qui ont été extraites des vidéos acoustiques. Cependant, à ce stade de notre développement, nous avons besoin d'une indication sur la capacité de discrimination des informations que nous extrayons. Les déformations successives de chaque individu sont extraites sous forme de matrices définies comme des cartes de déformations. Nous cherchons donc à les comparer entre elles et à leur attribuer des scores de similarité. Des méthodes de comparaison d'images telles que celle proposée par [Lowe, 2004] à travers le descripteur SIFT, qui fonctionne bien sur des images avec des valeurs d'intensité variables comme les nôtres [Karami et al., 2017], pourraient nous donner suffisamment d'indications pour produire ce score de similarité.

Dans ce chapitre, nous proposons donc une méthode qui combine le calcul de la déformation du corps des poissons à l'aide d'un modèle déformable, avec des outils de traitement d'image afin de fournir des scores de similarité permettant de discriminer les espèces.

5.1 Caractérisation du mode de nage : implémentation d'un modèle de déformation

Le modèle de déformation de maillage à gradient [Weng et al., 2006] que nous avons utilisé comme modèle de déformation fonctionne bien sur nos vidéos, ce qui nous permet de suivre de près les contours du corps du poisson à chaque détection dans le champ de vision de la caméra. Cependant, il faut noter que l'image binaire du corps du poisson donnée en entrée du modèle influence largement les résultats de la déformation. En cas de bruit ou d'occlusion dans l'image du corps du poisson, le modèle se déforme en fonction de ces artefacts, ce qui donne une déformation incorrecte du maillage du corps. Le calcul de la déformation qui en résulte peut donc également être faussé.

Calculées à l'aide des modèles déformables, les cartes de déformation produites par notre méthode sont difficilement lisibles et donc difficilement comparables. Leur post-traitement est donc une étape importante avant d'obtenir de précieuses informations. Les études sur le mode de nage proposées dans la littérature sont réalisées à partir de caméras optiques connues pour leur fréquence d'images de 24 images par seconde, permettant un bon suivi de la déformation du poisson tout au long de son déplacement. Avec des fréquences d'images variant de 5 à 8 images par seconde sur les vidéos enregistrées par les CA, nous pouvons capter les déformations mais de manière saccadée. L'interpolation de nos cartes de déformation le long de l'axe temporel permet donc de lisser les transitions et de mettre en évidence le processus de déformation auquel est soumis le corps du poisson. Les deux étapes suivantes de filtrage ont été réalisées afin de nettoyer ces cartes en se concentrant sur leurs caractéristiques principales afin de simplifier le processus de comparaison. : lissage par interpolation avec une gaussienne, et, reconstruction de l'image par rapport aux deux premières composantes de son ACP.

5.2 Cartes de déformation : un outil innovant pour la classification des espèces

Afin d'analyser la capacité discriminante des cartes de déformation ainsi générées, un échantillon des jeux de données MZC-ARIS et SEL-ARIS, enregistrés avec des caméras ARIS, est utilisé comme jeu de données de test. Au total, 31 passages de poissons sont analysés : 10 anguilles européennes, 10 saumons atlantiques, 5 silures, 5 poissons du groupe CYP et 1 du groupe PRED. Afin d'obtenir des indices de similarité entre les 31 poissons, leurs cartes de déformation sont comparées deux à deux à l'aide de la méthode SIFT.

L'analyse visuelle des cartes de déformations met en évidence des motifs de déformation intra-groupe et inter-groupe, ainsi qu'une grande variabilité intra-groupe des cartes de déformation. Cette variabilité influence les résultats des comparaisons des cartes de déformation et conduit à

l'identification des similarités entre individus de groupes différents, par exemple entre l'anguille européenne et le silure. Le mode de nage anguilliforme de ces deux espèces peut cependant expliquer ces résultats. La même conclusion peut être tirée concernant les similarités occasionnellement observées entre les poissons des groupes SAT et CYP qui ont tous deux un mode de nage subcarangiforme. Cependant, certaines similarités sont moins cohérentes même après inspection visuelle des cartes de déformation ce qui suggère qu'une autre méthode de calcul de la similitude pourrait être utilisée. Une méthode de recherche automatique des motifs caractéristiques de chaque espèce ou groupe d'espèces, pourrait par exemple être envisagée.

Cependant, les résultats sont globalement encourageants avec des similitudes observées entre les individus d'un même groupe d'espèces, par exemple pour les groupes EEL, SIL et CYP. Certains individus ne présentent cependant aucune similarité avec les individus de leur groupe ce qui peut s'expliquer notamment par des différences de trajectoire.

Cette observation suggère que des facteurs externes influencent les cartes de déformation et doivent donc être pris en compte dans leur analyse. Le débit, la direction et la trajectoire ainsi que la vitesse de nage des poissons influencent le comportement de nage et donc la déformation de leur corps. Pour permettre une analyse plus robuste des cartes de déformation calculées, il serait nécessaire d'étudier les passages des poissons avec différentes configurations de ces facteurs. Une étude en conditions contrôlées serait une première approche pour étudier les passages de différentes espèces. Par conséquent, pour développer davantage notre méthode sur la déformation du corps, il est essentiel d'augmenter le nombre de passages de poissons étudiés.

Cette analyse complète ainsi la méthode que nous proposons pour l'analyse des vidéos acoustiques et le processus d'identification des espèces. L'objectif est ici de tirer profit de la dernière information disponible et encore non analysée dans notre méthode, la déformation du corps du poisson au cours de son déplacement dans le champ de vision de la caméra. Ce travail reste une étude préliminaire avec des résultats encourageants qui nécessiteront des recherches ultérieures.

6 Discussion et conclusion

Mon travail de thèse propose ainsi une méthode permettant d'extraire plusieurs caractéristiques des passages de poissons disponibles au sein des vidéos acoustiques (caractéristiques morphologiques, de mouvement et du mode de nage) dans le but de réaliser une classification multi-espèces des cibles détectées. Si l'extraction des caractéristiques morphologiques et de mouvement a déjà été exploitée dans la littérature, la méthode proposée ici fonctionne sur deux modèles de CA différents (ARIS et BLUEVIEW), pour détecter et identifier les grands individus anguilliformes. Cependant, les outils que j'ai développés devraient théoriquement fonctionner

sur des types d'images similaires telles que celles enregistrées par les caméras acoustiques DIDSON et Oculus, la performance sur ces données devant néanmoins être évaluée. La méthode de description automatique du comportement de nage est quant à elle unique et innovante au regard de la littérature existante sur l'analyse des vidéos acoustiques. Les résultats préliminaires obtenus montrent une forte similarité pour les silures mais une forte variabilité intra-groupe pour les groupes anguilles et saumons. De nombreux paramètres peuvent affecter le comportement de nage des poissons et doivent donc être pris en compte pour exploiter correctement ces cartes de déformation dans le cadre d'une identification multi-espèces. Les conditions environnementales telles que le débit de l'eau [Kirk et al., 2015, Lenihan et al., 2020], le comportement du poisson (alimentation, chasse) et sa direction de migration (amont, aval) sont autant de facteurs qui influencent la façon dont le poisson va déformer son corps pour se déplacer. En plus des données extraites sur les vidéos acoustiques, les données environnementales ainsi que les connaissances écologiques sur les espèces devront donc être prises en compte dans le processus de classification, résumé à la Figure 94.

Pour réaliser une classification automatique des espèces à partir de toutes les caractéristiques collectées, le développement d'un modèle de classification est essentiel. Il permettrait de s'affranchir de tout paramètre choisi empiriquement, comme c'est le cas pour l'analyse morphologique et de mouvement proposée pour les espèces anguilliformes. Par conséquent, un modèle de classification doit être entraîné sur un premier ensemble de données et évalué sur un second afin d'évaluer sa capacité à labelliser correctement les données. Une fois entraîné, le modèle est capable d'effectuer une prédiction de classe sur de nouvelles données. Pour développer un tel modèle, il est nécessaire d'élargir nos jeux de données avec davantage d'individus, en particulier pour les groupes d'espèces pour lesquels un faible nombre de passages a été collecté. Idéalement, ces données devraient également être enregistrées dans des conditions environnementales variées et avec une bonne représentation de leurs différents comportements.

Un travail supplémentaire est nécessaire pour développer le modèle de classification, qui peut être conçu comme décrit ci-après. A chaque détection d'une cible sur les vidéos acoustiques, toutes les caractéristiques sont extraites et remplissent une matrice de taille NxM avec N le nombre de caractéristiques, égal pour chaque individu, et M le nombre de détections, qui peut varier d'un individu à l'autre. Une première étape consisterait donc à normaliser ces matrices à une taille similaire par des méthodes de réduction des données, par exemple [Joliffe and Morgan, 1992, Izenman, 2008, Wang and Zhang, 2013]. Les matrices ainsi obtenues pourraient alors être utilisées comme entrée d'un algorithme de classification. De nombreux algorithmes d'apprentissage automatique sont disponibles pour effectuer cette étape de classification [Ray, 2019]. Afin d'intégrer les connaissances écologiques dans la classification, l'intégration des connaissances a priori dans le processus de classification devrait être discutée [Lauer and Bloch, 2008, Abbate et al., 2010].

En conclusion, les résultats de notre méthode fournissent des informations quantitatives et qualitatives à partir de vidéos acoustiques qui sont nécessaires aux gestionnaires pour assurer la préservation des populations de poissons. Ce travail offre de nombreuses perspectives pour mieux comprendre le comportement des espèces étudiées, tout en offrant l'avantage d'être une méthode non intrusive. La méthode que j'ai développée constitue une avancée dans l'objectif de disposer d'un outil entièrement automatique, compatible avec une analyse en temps réels fournissant des informations à long terme sur les populations de poissons présentes sur un site de suivi. La conception et l'implémentation du modèle de classification sera l'étape suivante pour atteindre cet objectif.

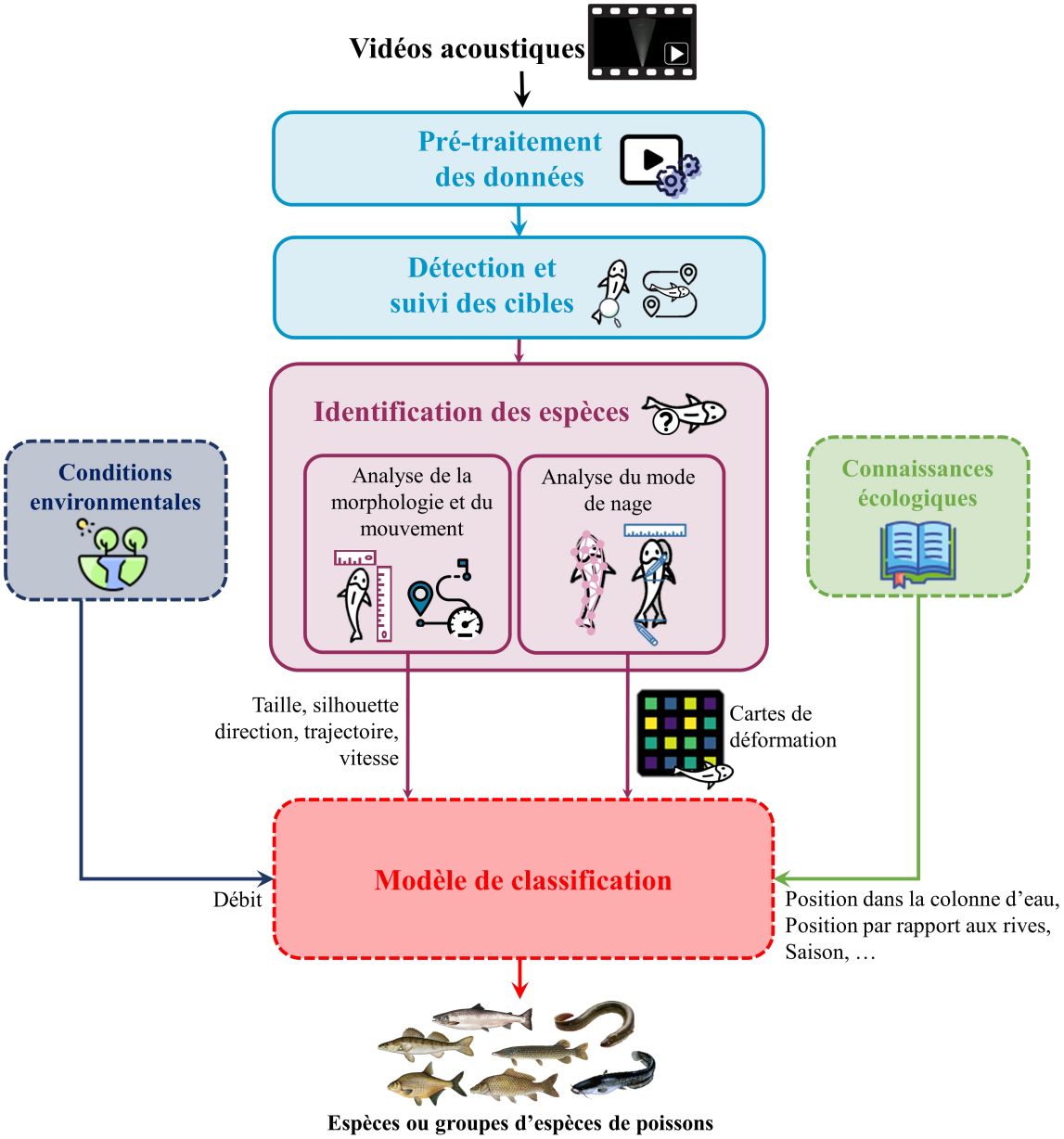


Figure 94 – Méthode proposée avec les différentes étapes du traitement affichées. Les étapes introduites dans ce travail (bleu clair et rose) et celles qui doivent encore être développées (en pointillé).

Appendix B. Comparison of two methods of automatic identification and counting of anguilliform species using acoustic cameras

As a prelude to the PhD work presented in this manuscript, a method based on Histogram of Oriented Gradient (HOG) and Support Vector Machine (SVM) classifier has been developed to automatically classify European eels in acoustic videos. The conference paper (GRETSI 2022) displayed below (in French) compares this HOG-SVM method with the one based on morphology and motion analysis proposed in Chapter 3. The results obtained show a much better performance for the second method (recall = 77.1%, precision = 81,9%, N = 170) than for the first (recall = 53.5%, precision = 54.8%, N = 170), reinforcing the choice of an approach based on the extraction of target characteristics via image processing tools.

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

AZENOR LE QUINIO[1], ALEXANDRE GIRARD[2], FABRICE ZAOUI[1], ERIC DE OLIVEIRA[1]

[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 visio*n. 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).

$$\forall\, w \in [1, W_1], \forall\, v \in [1, W_2],$$
$$\sigma_{w,v}^{N,*} = SVD^N(I_{i,j}^*)_{w*\frac{d_w}{2} \le i \le w*\frac{d_w}{2}+d_w,\, v*\frac{d_w}{2} \le j \le v*\frac{d_w}{2}+d_w} \quad (1)$$

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

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

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

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

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^{ième} valeur singulière de la fenêtre (w, v) ; $\sigma_{w,v}^{N,ref}$, la N^{ième} valeur singulière de la fenêtre (w, v) de référence ; S, la matrice des ratios ; M, le masque ; R, la matrice résultat des ratios.

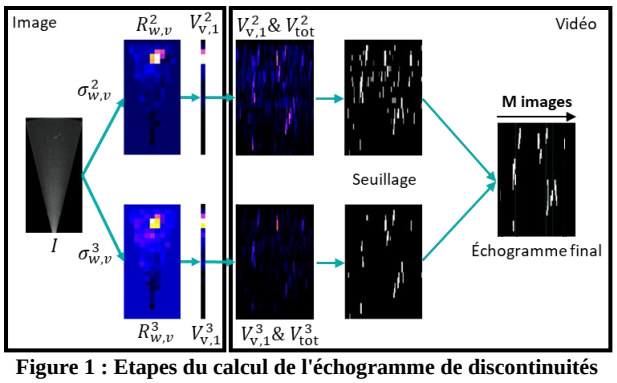


Figure 1 : Etapes du calcul de l'échogramme de discontinuités

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 2ème et 3ème 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.

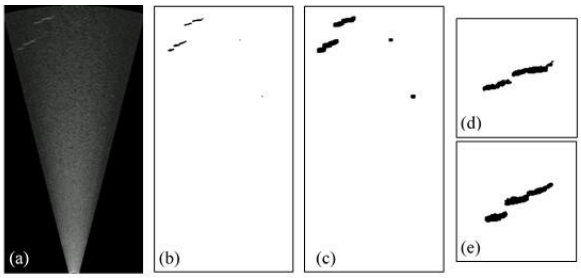


Figure 2 : a) Image traitée par filtre moyenneur ; b) Image binaire obtenue par soustraction d'arrière-plan ; c) Dilatation de l'image binaire ; d,e) Extraction des deux candidats

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.

A 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.

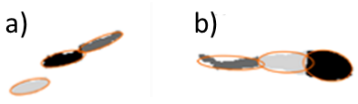


Figure 3 : Exemples de fragmentation et de calculs d'excentricité pour une espèce anguilliforme (a) et non anguilliforme (b)

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%.

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

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 ($R^2 \geq 0.91$).

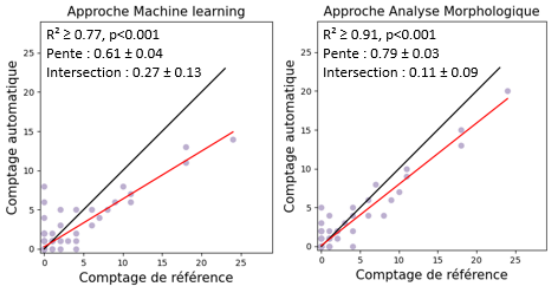


Figure 4 : Graphique des régressions linéaires entre comptage de référence et comptage automatique pour chaque approche

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.