INTER-FACULTY MASTER PROGRAM on

COMPLEX SYSTEMS and NETWORKS

SCHOOL of MATHEMATICS
SCHOOL of BIOLOGY
SCHOOL of GEOLOGY
SCHOOL of ECONOMICS

ARISTOTLE UNIVERSITY of THESSALONIKI

Master Thesis

Title:

Development of methods for the automated monitoring of fish in DIDSON (Dual-Frequency Identification Sonar) data

Triantafyllia-Maria Perivolioti

SUPERVISOR: Stefanos Sgardelis, Professor, AUTH
CO-SUPERVISOR: Michal Tušer, Research Fellow, Biology Centre CAS
CO-SUPERVISOR: Ioannis Antoniou, Professor, AUTH

Thessaloniki, December 2019

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ στα

ΠΟΛΥΠΛΟΚΑ ΣΥΣΤΗΜΑΤΑ και ΔΙΚΤΥΑ

ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΤΜΗΜΑ ΓΕΩΛΟΓΙΑΣ

ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Τίτλος Εργασίας:

Ανάπτυξη μεθόδων για την αυτοματοποιημένη παρακολούθηση ιχθύων σε δεδομένα DIDSON (Dual-Frequency Identification Sonar)

Τριανταφυλλιά-Μαρία Περιβολιώτη

ΕΠΙΒΛΕΠΩΝ: Στέφανος Σγαρδέλης, Καθηγητής, Α.Π.Θ.
ΣΥΝΕΠΙΒΛΕΠΩΝ: Michal Tušer, Συνεργάτης Ερευνητής, Biology Centre CAS
ΣΥΝΕΠΙΒΛΕΠΩΝ: Ιωάννης Αντωνίου, Καθηγητής, Α.Π.Θ.

Εγκρίθηκε από την Τριμελή Εξεταστική Επιτροπή την 18η Δεκεμβρίου 2019.

……	……	……
Σ. Σγαρδέλης	M. Tušer	I. Αντωνίου
Καθηγητής, Α.Π.Θ.	Συνεργάτης Ερευνητής, Biology Centre CAS	Καθηγητής, Α.Π.Θ.

Θεσσαλονίκη, Δεκέμβριος 2019

Copyright

……..
Τριανταφυλλιά-Μαρία Γ. Περιβολιώτη
Πτυχιούχος Βιολόγος Α.Π.Θ.

Abstract

The application of hydroacoustic techniques is well established in the field of Fisheries Science, as it is a methodology that can provide qualitative (e.g. behaviour, direction of movement) and quantitative (e.g. abundance, biomass) information of fish stock in high resolution and non-invasively collected. DIDSON is one of the most recent technological achievements in the field of Hydroacoustics. It is an acoustic camera that utilizes very high frequencies, thus having the capability to produce high quality images. As the DIDSON data can be properly converted to an image array, computer vision techniques can be used for processing the data. However, those methods almost invariably necessitate human intervention, in order to refine the final results and suitably fine-tune all processing parameters for an optimal result. The main aim of this study is to explore and identify the potential for automation of fish target detection from DIDSON data, combining techniques from the fields of computer vision and machine learning. Specifically, a workflow is proposed and validated, which is based on optical flow field calculation and the application of a genetic algorithm, in order to detect fish targets. Manual validation of the methodology indicated that at least half of the targets are successfully detected almost 60% of the time. The average neighborhood size (moving structure bounds) for the optical flow field calculation was estimated by the genetic algorithm to be around 12-14 pixels. The methodology provides promising results with minimal user intervention and can be extended to include additional parameters of the detection and tracking workflows.

Keywords: Hydroacoustics, acoustic imaging, image classification, fish detection

Contents

Σύνοψη

Τα τελευταία χρόνια τα εσωτερικά ύδατα, και ιδιαίτερα τα ποτάμια, αντιμετωπίζουν βαθμιαία υποβάθμιση της ποιότητας τους λόγω των ανθρωπογενών δραστηριοτήτων. Μία από τις σημαντικότερες προκλήσεις είναι η επίτευξη της διατήρησης της λειτουργίας των οικοσυστημάτων γλυκού νερού και η μείωση του ρυθμού απώλειας της βιοποικιλότητας. Για τον σκοπό αυτόν καθίσταται αναγκαία η υιοθέτηση νέων στρατηγικών για την παρακολούθηση της κατάστασης του υδάτινου περιβάλλοντος, οι οποίες θα είναι "φιλικές" προς τα υπό παρακολούθηση οικοσυστήματα. Συχνά, οι μέθοδοι παρακολούθησης των ιχθύων σε ποτάμια απαιτούν τη λήψη δειγμάτων, με αποτέλεσμα τη μείωση της βιομάζας των ιχθυοπληθυσμών. Στο πλαίσιο αυτό, η υδροακουστική αποτελεί μη παρεμβατική μέθοδο με δυνατότητα συνεχούς συλλογής δεδομένων με μεγάλη χωρική και χρονική ανάλυση. Άλλα πλεονεκτήματα της υδροακουστικής είναι α) η εύκολη και γρήγορη απεικόνιση των συλλεχθέντων δεδομένων, β) η δυνατότητα σύγκρισης και αξιολόγησης διαχρονικών δεδομένων και γ) η ψηφιακή μορφή των δεδομένων, η οποία επιτρέπει την ευκολότερη ανάκτηση και την ταχύτερη ανάλυση, μειώνοντας το κόστος, τον χρόνο επεξεργασίας μεγάλου όγκου δεδομένων.

Η υδροακουστική βασίζεται σε όργανα, τα οποία μεταδίδουν και λαμβάνουν ηχητικά κύματα. Με τον τρόπο αυτόν μπορούν να ανιχνεύσουν ψάρια ή άλλα αντικείμενα αποτελεσματικότερα από τα οπτικά συστήματα, δεδομένου ότι ο ήχος διαπερνά το νερό αποτελεσματικότερα από ό,τι το φως. Οι εφαρμογές της υδροακουστικής, όσον αφορά στο πεδίο της Αλιευτικής Έρευνας, άπτονται της παρακολούθησης της συμπεριφοράς, της αφθονίας και της κατανομής των ιχθύων ή άλλων υδρόβιων οργανισμών (π.χ. πλαγκτόν), καθώς και της παρακολούθησης και αξιολόγησης των επιπτώσεων των κοινών δειγματοληπτικών εργαλείων στους ιχθυοπληθυσμούς. Επειδή η υδροακουστική είναι μια ευρέως διαδεδομένη μέθοδος στον τομέα της Αλιευτικής Έρευνας, έχει κατασκευαστεί πλήθος συσκευών που βασίζονται στην εκπομπή μιας ηχητικής συχνότητας και την ανίχνευση της ανάκλασής της από σώματα διαφορετικής πυκνότητας.

Η ακουστική κάμερα DIDSON είναι ένα ηχοβολιστικό υψηλής συχνότητας ικανό να παράγει εικόνες υψηλής ανάλυσης μέσω ενός μοναδικού προσαρμοσμένου συστήματος ακουστικών φακών. Η ακουστική κάμερα DIDSON, που αναπτύχθηκε από

το Πανεπιστήμιο της Ουάσιγκτον, ΗΠΑ (Applied Physics Lab), μπορεί να δημιουργήσει βίντεο υψηλής ανάλυσης. Ως αποτέλεσμα, η παρακολούθηση των ιχθύων μπορεί να επιτευχθεί σε πολύ καλύτερα επίπεδα λεπτομέρειας από αυτά που μπορούν να προσφέρουν οι κοινές τεχνολογίες ηχοβολιστικών. Επιπλέον, είναι δυνατή η ανίχνευση και η οπτική αναγνώριση του θορύβου (π.χ. φυσαλίδες, αιωρούμενη ύλη κ.λπ.).

Η ακουστική κάμερα DIDSON διαθέτει δύο συχνότητες λειτουργίας, 1.8 MHz (η λειτουργία υψηλής συχνότητας) και 1.1 MHz (λειτουργία χαμηλής συχνότητας). Στη λειτουργία υψηλής συχνότητας, η συνολική ακτίνα αποτελείται από 96 υπο-ακτίνες, ενώ η απόσταση ακτινοβόλησης μπορεί να ρυθμιστεί από 12 έως και 15 μέτρα. Στη λειτουργία χαμηλής συχνότητας, η συνολική ακτίνα αποτελείται από 48 υπο-ακτίνες και η απόσταση ακτινοβόλησης μπορεί να ρυθμιστεί έως και 40 μέτρα. Το συνολικώς ακτινοβολούμενο πεδίο έχει εύρος 29° x 14° (οριζόντια x κατακόρυφη διάσταση) και στις δύο καταστάσεις λειτουργίας. Το μήκος του παλμού (διάρκεια μεταξύ διαδοχικών παλμών) εξαρτάται από τις ρυθμίσεις απόστασης και κυμαίνεται από 4.5 έως και 144 ms. Ως εκ τούτου, υποστηρίζονται ρυθμοί καταγραφής έως και 21 καρέ ανά δευτερόλεπτο.

Τα οικοσυστήματα γλυκού νερού είναι ετερογενή και οι περιβαλλοντικοί παράγοντες ποικίλλουν μεταξύ των υδροακουστικών καταγραφών. Ως εκ τούτου, απαιτούνται αποτελεσματικές μέθοδοι ανάλυσης για την επεξεργασία δεδομένων DIDSON. Έχουν αναπτυχθεί αρκετά λογισμικά ανάλυσης, τα οποία μπορούν να ανιχνεύσουν, να καταμετρήσουν και να παρακολουθήσουν τη συμπεριφορά των ιχθύων από δεδομένα DIDSON. Συγκεκριμένα, τα λογισμικά DIDSON (Sound Metrics Corp.), ECHOVIEW (Myriax Pty Ltd, Hobart, TAS, Australia) και Sonar5-Pro (Lindem Data Acquisition, Όσλο, Norway) είναι τα πιο ευρέως χρησιμοποιούμενα λογισμικά για την ανάλυση των δεδομένων DIDSON. Επίσης, δεδομένου ότι τα δεδομένα DIDSON μπορούν να μετατραπούν σε υψηλής ανάλυσης βίντεο, οι περισσότερες σύγχρονες μεθοδολογίες χρησιμοποιούν τεχνικές επεξεργασίας εικόνας. Ωστόσο, οι υφιστάμενες μέθοδοι ανάλυσης των δεδομένων σχεδόν πάντοτε απαιτούν ανθρώπινη παρέμβαση, προκειμένου να τελειοποιήσουν τα τελικά αποτελέσματα με την απομάκρυνση αντικειμένων ή άλλων τυχαίων αποτελεσμάτων. Επιπλέον, ενώ η αυτοματοποίηση επιτυγχάνεται πρακτικά, οι διαδικασίες βασίζονται συνήθως σε εκτεταμένη παραμετροποίηση και απαιτούν σημαντική εμπειρία για την κατάλληλη τελειοποίηση όλων των παραμέτρων επεξεργασίας για ένα βέλτιστο αποτέλεσμα.

Ο πρωταρχικός σκοπός της παρούσης εργασίας είναι να διερευνήσει και να βελτιώσει τις τρέχουσες μεθοδολογίες για τον αυτοματοποιημένο εντοπισμό και παρακολούθηση ιχθύων από τα δεδομένα του DIDSON. Στο πλαίσιο της βελτίωσης των υφιστάμενων μεθοδολογιών ανίχνευσης και παρακολούθησης ιχθύων καθορίζονται δύο βασικοί στόχοι:

- Ο πρώτος στόχος είναι η πρόταση και η διερεύνηση της αποτελεσματικότητας μιας σχετικά πρόσφατης τεχνικής ανίχνευσης κίνησης, αυτής της οπτικής ροής, για την παρακολούθηση και την ανίχνευση ιχθύων – στόχων σε δεδομένα DIDSON. Αυτό θα επιτευχθεί με την οργάνωση και υιοθέτηση μιας συγκεκριμένης ροής εργασιών, η οποία θα περιέχει ενσωματωμένους σε αυτήν υπολογισμούς οπτικής ροής, προσδιορίζοντας, έτσι, όλες τις παραμέτρους που επηρεάζουν με την ποιότητα του αποτελέσματος.
- Ο δεύτερος στόχος είναι η πρόταση και η υιοθέτηση μιας μεθοδολογίας για την αυτοματοποίηση της διαδικασίας επιλογής συγκεκριμένων παραμέτρων στην υφιστάμενη ροή εργασίας. Αυτό θα επιτευχθεί μέσω του συνδυασμού της προαναφερθείσας ροής εργασιών με μια μέθοδο αναπροσαρμογής παραμέτρων, συγκεκριμένα με έναν γενετικό αλγόριθμο.

Το προτεινόμενο ερευνητικό πλαίσιο θα εφαρμοστεί σε ένα σύνολο δεδομένων DIDSON, το οποίο αποκτήθηκε από ηχητικές καταγραφές στον ποταμό Μολδάβα.

Στην παρούσα εργασία, χρησιμοποιήθηκαν δεδομένα υδροακουστικής, τα οποία συλλέχθηκαν στο πλαίσιο του έργου "Coexistence of human and pearl mussel *Margaritifera margaritifera* in the Vltava River" με σκοπό την παρακολούθηση της μετανάστευσης των ιχθύων στον ποταμό Μολδάβα, Δημοκρατία της Τσεχίας. Συγκεκριμένα, έγιναν καταγραφές με τη χρήση ακουστικής κάμερας DIDSON για την παρακολούθηση της διέλευσης των ιχθύων από τον ταμιευτήρα Lipno στον ποταμό Μολδάβα, στην περιοχή του Εθνικού Πάρκου Šumava (48 ° 48.52115'N, 13 ° 56.77817'E). Η παρακολούθηση της διέλευσης των ιχθύων πραγματοποιήθηκε για δύο περιόδους ωοτοκίας (5/4 / 2014-9 / 5/2014 και 18/3 / 2015-18 / 6/2015). Κατά τη διάρκεια αυτών των περιόδων η ακουστική καταγραφή ήταν σχεδόν συνεχής. Η ακουστική κάμερα DIDSON χρησιμοποιήθηκε στη λειτουργία υψηλής συχνότητας για τη λήψη καλύτερης ποιότητας δεδομένων.

Η οπτική ροή έχει οριστεί ως η κατανομή των ταχυτήτων της κίνησης των μοτίβων φωτεινότητας σε μια εικόνα. Θεωρώντας τη σημειακή φωτεινότητα ως το κύριο

συστατικό της οπτικής αναπαράστασης, η οπτική ροή μπορεί να περιγραφεί ως το πεδίο της ταχύτητας (ή της στιγμιαίας μετατόπισης) της φωτεινότητας. Λαμβάνοντας υπόψη την τυπική οπτική αναπαράσταση μιας εικόνας ως συστοιχία στοιχείων εικόνας (εικονοστοιχεία) με καθορισμένες τιμές φωτεινότητας B (x, y) και δεδομένου ότι η εικόνα είναι μια αναπαράσταση συνεκτικών και χρονικά διαδοχικών καταστάσεων, τα αποτελέσματα της κίνησης που αποδίδονται με την εισαγωγή της χρονικής μεταβλητής στο πεδίο φωτεινότητας της εικόνας, B (x, y, t). Ο τελικός στόχος της μεθόδου της οπτικής ροής είναι η εξαγωγή των πιθανών μετατοπίσεων αντικειμένων σε μια ακολουθία εικόνων, με βάση την οπτική αναπαράσταση του χρονικά μεταβαλλόμενου πεδίου φωτεινότητας.

Για να καταστεί δυνατός ο προσδιορισμός της μετατόπισης με βάση τη μέθοδο της οπτικής ροής θα πρέπει να ληφθούν υπόψη ορισμένες παραδοχές. Η πιο σημαντική παραδοχή είναι ότι η επίδραση της φυσικής κίνησης μεταφράζεται ως μια μετατόπιση της έντασης ενός εικονοστοιχείου σε ένα διαφορετικό, συνήθως γειτονικό, εικονοστοιχείο. Αυτή η παραδοχή ονομάζεται περιορισμός της σταθερότητας της φωτεινότητας και πρακτικά συνεπάγεται την παραδοχή, ότι η φωτεινότητα όλων των αντικειμένων που ανήκουν στην εικόνα είναι σταθερή.

Έχουν αναπτυχθεί διάφορες διαφορικές μέθοδοι για τον υπολογισμό της οπτικής ροής. Ένας από τους πιο πρόσφατους αλγόριθμους υπολογισμού οπτικής ροής βασίζεται στις τεχνικές του Gunnar Farnebäck για εκτίμηση της κίνησης χρησιμοποιώντας τετραγωνικά πολυώνυμα. Συγκεκριμένα, χρησιμοποιώντας ζεύγη εικόνων, κάθε υπό-περιοχή της εκάστοτε εικόνας του ζεύγους προσεγγίζεται χρησιμοποιώντας τετραγωνικά πολυώνυμα. Το πεδίο μετατόπισης, στη συνέχεια, υπολογίζεται προσαρμόζοντας κατάλληλα ένα γενικό μοντέλο πολυωνυμικού μετασχηματισμού. Η καινοτομία των αλγορίθμων εκτίμησης κίνησης Farnebäck ήταν η χρήση τανυστών προσανατολισμού, καθώς και η συμπερίληψη παραμετρικών μοντέλων κίνησης. Οι υπολογισμοί που έγιναν στο πλαίσιο αυτής της εργασίας πραγματοποιήθηκαν με τη χρήση του λογισμικού MATLAB® και του αλγόριθμου εκτίμησης της οπτικής ροής Farnebäck.

Ο γενετικός αλγόριθμος είναι η μαθηματική έννοια, η οποία μπορεί να θεωρηθεί ως το αλγοριθμικό ομόλογο της εξέλιξης. Ένα πρόβλημα, το οποίο έχει μοντελοποιηθεί ως ένα παραμετροποιημένο σύστημα μπορεί να έχει ως λύσεις διάφορα σύνολα παραμέτρων, τα οποία μπορούν να τηρούν ποικίλους περιορισμούς. Οι γενετικοί αλγόριθμοι είναι, ως εκ τούτου, μια ομάδα μεθόδων που στοχεύει στην εύρεση βέλτιστων

προσεγγίσεων σε αυτά τα σύνολα λύσεων, δεδομένου ενός αριθμού αρχικών συνόλων και λαμβάνοντας υπόψη ένα κριτήριο αξιολόγησης. Υπάρχουν δύο γενικές κατηγορίες κριτηρίων αξιολόγησης, τα κριτήρια καταλληλότητας και τα κριτήρια ποινής. Είναι σημαντικό να αναφερθεί, ότι οι λύσεις που προκύπτουν από τους γενετικούς αλγόριθμους δεν είναι απαραιτήτως οι βέλτιστες λύσεις στα αντίστοιχα προβλήματα. Στην πραγματικότητα, επειδή οι γενετικοί αλγόριθμοι μιμούνται τις εξελικτικές διεργασίες, δεν είναι πάντοτε δυνατό να ερμηνευθούν λογικά σε μαθηματικό επίπεδο.

Απαραίτητη προϋπόθεση για τη χρήση του γενετικού αλγορίθμου στην επίλυση ενός προβλήματος, είναι η δυνατότητα αξιολόγησης των λύσεων, ανεξαρτήτως της δυνατότητας επίλυσης του προβλήματος. Τα βασικά βήματα ενός γενετικού αλγορίθμου είναι, ενδεικτικά, τα εξής :

- Κάθε υποψήφια λύση αναπαρίσταται με μια συμβολοσειρά, η οποία αναφέρεται ως χρωμόσωμα.
- Παρέχεται ένα σύνολο αρχικών λύσεων (ακόμη και τυχαίων).
- Εφαρμογή σειράς γενετικών φορέων και ανασυνδυασμός των λύσεων.
- Μετάβαση στην επόμενη γενιά (η οποία διαμορφώνεται ως επαναληπτική διαδικασία) εφαρμόζοντας τη διαδικασία της επιλογής.

Αυτά τα βήματα είναι όμοια με τη βιολογική εξέλιξη και, με το πέρασμα των γενιών, οδηγούν στη συνεχή βελτίωση των συνόλων λύσεων.

Μια σημαντική εφαρμογή γενετικών αλγορίθμων, η οποία σχετίζεται άμεσα με την παρούσα εργασία, είναι η χρήση τους στην επίλυση μη-γραμμικών μεικτών προβλημάτων ακεραίων με περιορισμένο πεδίο τιμών. Επειδή οι γενετικοί αλγόριθμοι δεν απαιτούν πρόσθετες πληροφορίες σχετικά με τα προβλήματα ή την απόκριση του προβλήματος σε διάφορες λύσεις, είναι κατάλληλοι για προβλήματα με ασαφείς ορισμούς ή/και ελλιπείς προδιαγραφές. Τα προβλήματα ακεραίων είναι εκείνα, στα οποία το σύνολο λύσεων είναι πιθανώς οριοθετημένο και ένα υποσύνολο των παραμέτρων λαμβάνει τιμές από το σύνολο των ακεραίων. Επειδή ορισμένες από τις παραμέτρους που σχετίζονται με τον προσδιορισμό της οπτικής ροής είναι εξ ορισμού ακέραιες τιμές (π.χ. το μέγεθος του φίλτρου), οι γενετικοί αλγόριθμοι είναι κατάλληλοι για τον προσδιορισμό των βέλτιστων τοπικών λύσεων.

Η ροή εργασιών που ακολουθήθηκε στην παρούσα εργασία αποτελείται από ένα σταθερό τμήμα και ένα επαναληπτικό μέρος, το οποίο βασίζεται σε μια προκαθορισμένη τιμή επαναλήψεων, η οποία ορίζεται από τον χρήστη. Επιπλέον, ένα μέρος της

επαναληπτικής διαδικασίας βασίζεται στην επιλογή ενός αριθμού N συνόλων παραμέτρων, οι οποίες επανακαθορίζονται κατά τον υπολογισμό των αντίστοιχων πεδίων οπτικής ροής. Συγκεκριμένα, προκειμένου να εξαχθεί η επιθυμητή πληροφορία από τα δεδομένα DIDSON, ακολουθείται η παρακάτω διαδικασία:

1. Εξαγωγή των δεδομένων από το αντίστοιχο ψηφιακό αρχείο DIDSON. Κατά τη φάση αυτή, τα δεδομένα μετατρέπονται από την ψηφιακή μορφή τους σε πίνακες 2-διαστάσεων με τιμές ανακλώμενης έντασης, οι οποίοι ερμηνεύονται ως εικόνες στην κλίμακα του γκρι. Η αλληλουχία των εικόνων παράγει ένα αρχικό βίντεο, το οποίο μπορεί να χρησιμοποιηθεί για περαιτέρω επεξεργασία. Στο βήμα αυτό, οι τιμές έντασης μετατρέπονται έπειτα και από βαθμονόμηση σε μεγέθη κατάλληλα για περαιτέρω χρήση (τιμές Sv – Volume Backscattering Coefficient), ενώ αποκαθίσταται εικονικά η γεωμετρία της αρχικής λήψης, σύμφωνα με την οποία το ακουστικό πεδίο θέασης διευρύνεται με την απομάκρυνση από τον πομπό του ηχοβολιστικού μηχανήματος.
2. Προεπεξεργασία του αποτελέσματος για την απομάκρυνση του θορύβου. Συγκεκριμένα, εφαρμόζεται φίλτρο χρονικής εξομάλυνσης, το οποίο χρησιμοποιεί κινητό παράθυρο μεγέθους 8 καρέ (= 1 δευτερόλεπτο με βάση τον συγκεκριμένο ρυθμό καταγραφής των χρησιμοποιηθέντων δεδομένων).
3. Μοντελοποίηση του "παρασκηνίου" και του "προσκηνίου" και διαχωρισμός μεταξύ των κινητών και των ακίνητων τμημάτων για κάθε εικόνα-καρέ της ακολουθίας. Συγκεκριμένα, τα ακίνητα μέρη μοντελοποιούνται ως μία χρονική υστέρηση 8 καρέ, δηλαδή για κάθε εικόνα-καρέ, τα ακίνητα μέρη αντιστοιχίζονται ως η εικόνα-καρέ, η οποία προηγείται αυτής χρονικά κατά 8 καρέ. Η αφαίρεση μεταξύ των δύο οδηγεί στην εξαγωγή των κινητών τμημάτων για κάθε εικόνα της ακολουθίας. Λόγω του γεγονότος, ότι τα 8 πρώτα καρέ δεν έχουν ομόλογο καρέ "παρασκηνίου", αυτά εξαιρούνται από την τελική ακολουθία εικόνων.
4. Περιορισμός των περιοχών κίνησης των στόχων και διαχωρισμός από τα ακίνητα μέρη (παρασκήνιο) με χρήση αλγορίθμου υπολογισμού οπτικής ροής. Στο βήμα αυτό υπολογίζονται οι μετατοπίσεις με χρήση αλγορίθμου οπτικής ροής, και εφαρμόζεται αλγόριθμος εντοπισμού διαχωριστικής οριακής τιμής (thresholding) για τον προσδιορισμό εξομαλυμένων γενικών περιοχών, όπου κινούνται οι

στόχοι. Ως αλγόριθμος εντοπισμού οριακής διαχωριστικής τιμής χρησιμοποιείται η μέθοδος του Otsu. Το αποτέλεσμα αυτού του βήματος είναι μία πρώτη "μάσκα", η οποία διαχωρίζει τις γενικές περιοχές όπου εντοπίζεται κίνηση, από το υπόλοιπο ακίνητο τμήμα της εικόνας.

5. Περαιτέρω διαχωρισμός και αποσαφήνιση των στόχων από τα κινητά μέρη. Στο συγκεκριμένο βήμα, στις εικόνες εφαρμόζεται εκ νέου αλγόριθμος εντοπισμού οριακής διαχωριστικής τιμής (threshold) αναλύοντας μόνον τα απομονωμένα pixel των γενικών περιοχών, τα οποία περιλαμβάνονται στην μάσκα του προηγούμενου βήματος. Για το βήμα αυτό απαιτούνται ως παράμετροι εισόδου το μέγεθος της περιοχής φίλτρου (filter size) και το μέγεθος της περιοχής ανάλυσης (neighborhood size), οι οποίες είναι χαρακτηριστικές της μεθόδου υπολογισμού οπτικής ροής, η οποία χρησιμοποιήθηκε, συγκεκριμένα, του αλγορίθμου του Farnebäck. Ο προσδιορισμός των παραμέτρων έλαβε χώρα ακολουθώντας το επόμενο βήμα.
6. Εφαρμόζεται γενετικός αλγόριθμος για τον προσδιορισμό των υπό συνθήκες βέλτιστων τιμών για τις προαναφερθείσες παραμέτρους. Η συνθήκη βελτιστοποίησης εκφράζεται ως μια συνάρτηση αξιολόγησης της τελικής λύσης, δηλαδή της τελικώς υπολογισμένης μάσκας των στόχων. Η συνάρτηση αξιολόγησης εκφράζεται ως κριτήριο ποινής, με αποτέλεσμα να επιδιώκεται ο εντοπισμός λύσεων, οι οποίες ελαχιστοποιούν τη συνάρτηση αυτή. Για σκοπούς σύγκρισης εφαρμόστηκαν δύο διαφορετικές συναρτήσεις ποινής. Η μία από αυτές υπολογίζει ως αποτέλεσμα της αξιολόγησης το μέσο πλήθος από εικονοστοιχεία της μάσκας ανά εικόνα. Η δεύτερη εφαρμόζει αντίστοιχη αριθμητική ποινή για πολύ μεγάλα ή για πολύ μικρά εντοπισμένα αντικείμενα. Για αντικείμενα με πλήθος εικονοστοιχείων ανάμεσα στα όρια εφαρμόζεται ποινή αντιστρόφως ανάλογη του πλήθους.

Μετά το πέρας της αλγοριθμικής διαδικασίας υπολογισμών, η τελική μάσκα με τη βέλτιστη αξιολόγηση (ελάχιστη ποινή) επιλέγεται ως το τελικό αποτέλεσμα της διαδικασίας.

Τα δεδομένα που χρησιμοποιήθηκαν για την ανάλυση ανακατασκευάστηκαν σε μια αλληλουχία εικόνων μεγέθους [717 pixel * 400 pixel] και χρησιμοποιήθηκε ένα υποσύνολο 1000 εικόνων για περαιτέρω ανάλυση. Ο γενετικός αλγόριθμος ρυθμίστηκε για να εξελίξει 5 γενιές 6 μεμονωμένων λύσεων.

Τα αποτελέσματα έδειξαν ότι οι μεμονωμένες λύσεις ανήκουν σε δύο ομάδες, οι οποίες χαρακτηρίζονται από πολύ μεγάλη διαφορά στις εκτιμώμενες ποινές, με πολλαπλές τάξεις μεγέθους, παρόμοιες με τη διαφορά στις εφαρμοζόμενες ποινές. Εκείνες που εμφανίζουν πολύ μεγάλες τιμές περιέχουν ένα ή περισσότερα αντικείμενα μεγάλου μεγέθους στις μάσκες. Οι επιμέρους λύσεις που παρουσιάζουν μικρή ποινή διατηρούνται μεταξύ των γενεών και, τελικά, η μέση ποινή ανά γενιά παρουσιάζει μείωση.

Δύο ήταν οι βέλτιστες λύσεις που προέκυψαν, ανάλογα με τη συνάρτηση ποινής. Η λύση 5-12 (μέγεθος φίλτρου 5 εικονοστοιχείων και μέγεθος γειτονιάς 12 εικονοστοιχείων) από την εφαρμογή του πρώτου κριτηρίου ποινής, και η λύση 33-14 (μέγεθος φίλτρου 33 εικονοστοιχείων και μέγεθος γειτονιάς 14 εικονοστοιχείων) από την εφαρμογή του δεύτερου κριτηρίου ποινής. Η λύση 33-14 έχει γενικά υψηλότερο ποσοστό επιτυχίας ανίχνευσης πραγματικών στόχων σε σύγκριση με τη λύση 5-12. Ωστόσο, η λύση 33-14 έχει περισσότερες εσφαλμένες ανιχνεύσεις στόχων ανά εικόνα. Η λύση 33-14 οδηγεί στο σχηματισμό καλύτερα οριοθετημένης μάσκας. Αυτό το αποτέλεσμα εξηγείται από τη μέθοδο ανίχνευσης, η οποία είχε οριστεί για να αποκλείσει τις ανιχνεύσεις στόχων συνολικού μεγέθους <50 εικονοστοιχεία. Από την αριθμητική σύγκριση της απόδοσης των δύο λύσεων για όλες τις εικόνες προκύπτει ότι η λύση 33-14 έχει εμφανές πλεονέκτημα απόδοσης σε σωστές ανιχνεύσεις.

Η παρούσα εργασία αποτελεί μια προσπάθεια διερεύνησης της δυνατότητας ανάπτυξης μιας αυτοματοποιημένης, παραμετρικά βελτιστοποιημένης ροής εργασίας, που περιλαμβάνει την ανίχνευση κίνησης με τη μέθοδο της οπτικής ροής, για την βελτίωση της ανίχνευσης ιχθύων σε δεδομένα DIDSON. Η προτεινόμενη μεθοδολογία στοχεύει στην ελαχιστοποίηση της παρέμβασης του χρήστη με τη χρήση αυτοματοποιημένων και αυτορρυθμιζόμενων αλγοριθμικών διαδικασιών, όπως αυτή του γενετικού αλγορίθμου.

Οι βασικές παράμετροι του μεγέθους του φίλτρου και του μεγέθους της γειτονιάς για τον υπολογισμό του πεδίου οπτικής ροής από την ακολουθία εικόνων DIDSON επηρεάζουν σημαντικά την τελικά υπολογισμένη μάσκα. Η λύση [5, 12] φαίνεται ότι ελαχιστοποιεί υπό όρους τα συνολικά εικονοστοιχεία που καλύπτονται για κάθε εικόνα, ενώ η λύση [33,14] ελαχιστοποιεί τον συνολικό αριθμό διακριτών συνδεδεμένων στοιχείων. Η συνολική αλγοριθμική διαδικασία οδηγεί επιτυχώς σε μια τοπικά βέλτιστη μάσκα από τα δεδομένα DIDSON, οδηγώντας σε ικανοποιητικά αποτελέσματα με

ελάχιστη παρέμβαση του χρήστη. Ωστόσο, ο αλγόριθμος εξαρτάται σημαντικά από την επιλογή του κατάλληλου κριτήριου ποινής.

Μια άλλη παρατήρηση που αξίζει να αναφερθεί είναι ότι τα δύο διαφορετικά κριτήρια ποινής οδηγούν σε σύγκλιση σχεδόν στο ίδιο μέγεθος γειτονιάς, ήτοι 12-14 εικονοστοιχεία. Αυτό είναι το μέσο μέγεθος της περιοχής, όπου μπορεί να ανιχνευθεί βέλτιστα η κίνηση, η οποία με τη σειρά της μπορεί να ερμηνευθεί ως το μεγαλύτερο δυνητικό μέγεθος των παρατηρούμενων μετακινούμενων δομών.

Ενώ η αλγοριθμική διαδικασία χρησιμοποιήθηκε με επαρκή επιτυχία σε δύο διαφορετικά σύνολα δεδομένων, τα αποτελέσματα χρειάζονται επικύρωση συγκρίνοντας με εναλλακτικές μεθοδολογίες. Ένα από τα σημαντικότερα μειονεκτήματα της διαδικασίας είναι ο σχετικά μεγάλος χρόνος που απαιτείται για τον υπολογισμό του αλγορίθμου υπολογισμού οπτικού πεδίου ροής.

Acknowledgements

At this point, I would like to thank all the people that have contributed to this research in their own particular ways.

I owe my deepest gratitude to my supervisor, Prof. Stefanos Sgardelis, who has supported me throughout my thesis with his patience, knowledge, guidance and support.

I would like to express my sincere gratitude to Michal Tušer, Research Fellow at Biology Centre of the Czech Academy of Sciences, Institute of Hydrobiology, for all the knowledge he shared with me, but also for his encouragement, his patience, his contagious enthusiasm and friendship. His guidance helped me in all the time of research and writing of this thesis. Without his guidance and persistent help this master thesis would not have been possible.

It is with immense gratitude that I acknowledge the support and help of Prof. Ioannis Antoniou, who introduced me to the field of Complex Systems and Networks. His dynamism, vision, sincerity and motivation have deeply inspired me. Thank you for investing time and providing interesting and valuable feedback.

I would like to thank Prof. RNDr. Jan Kubečka, CSc., the Director of the Institute of Hydrobiology at Biology Centre CAS, who provided me an opportunity to join his team. Also, I am very grateful to the members of the FishEcU (Fish Ecology Unit of the Department of Fish and Zooplankton Ecology, Institute of Hydrobiology CAS) for their feedback, cooperation and of course friendship. Especially, my great appreciation and true thanks is for Dr. Jaroslava Frouzova for sharing with me her knowledge on hydroacoustics.

I would like to thank my Ph.D supervisor, Dimitra Bobori, and Dr. Antonios Mouratidis, Assistant Professor, School of Geology, for their patience and support in overcoming numerous obstacles I have been facing through my research and in developing myself as a researcher in the best possible way.

My sincere thanks to Dimitrios Terzopoulos for the continuous encouragement and for accepting nothing less than excellence from me.

I am grateful to all the academic staff members of the Interfaculty Master Program ‘Complex Systems and Networks’ for offering me their scientific and technical experiences.

Last but not the least, I would like to thank my family supporting me spiritually throughout writing this thesis and my life in general.

Abbreviations

DIDSON	**D**ual-frequency **ID**entification **SON**ar
ARIS	**A**daptive **R**esolution **I**maging **S**onar
Sonar	**So**und **N**avigation **A**nd **R**anging

1. Introduction

As inland waters are faced with a gradual deterioration of water quality, it is imperative to develop new methods and strategies to monitor and reverse the condition of the aquatic environment. Fisheries science is generally concerned with both the biological and physical aspects and focuses on cross-manipulating these aspects in order to achieve regulation of the corresponding processes. To be effective, this field strongly depends on the availability of fast and detailed analytical monitoring tools. One such tool, which has the potential to enhance the methodologies of Fisheries science is hydroacoustics, which, unlike traditional methods, offers high-resolution data, containing dense spatial information. This information can quickly and easily be visualized, evaluated and cross-compared across water bodies or periods (e.g. seasonal or annual variations). Also, hydroacoustic data are collected in a digital format, enabling easier retrieval and very efficient potential for analysis, and reducing cost, time and effort for the processing of large volumes of data. The combination of hydroacoustics with capture methods, in order to establish ground truth information comprises a great tool for tackling the key ecohydrological concerns of how fish populations interact with their habitats, by mapping fish spatial distribution, biomass and behavioral aspect responses under habitat changes (Godlewska et al., 2002; Winfield, 2004).

1.1. Principles of Fisheries Acoustics

Active remote-sensing instruments based on the transmission and reception of sound waves (acoustic instruments) can detect fish, or other targets, at a far greater range compared to visual systems, because sound penetrates water significantly more than light. As a result, acoustic instruments have become a commodity of fishing vessels, as they can determine fish and seabed location as well. Applications based on acoustics have extensively pervaded the field of fisheries science as a tool for the assessment of behaviour, abundance and distribution of fish or other aquatic organisms (e.g. plankton), as well as the performance monitoring and assessment of sampling gears (Walsh et al., 2001; Fernandes et al., 2002; Rakowitz et al., 2012).

The fundamental tool used in acoustics applications in fisheries science is a scientific echosounder. An echosounder's transducer converts an electrical signal to an

acoustic pulse, which is termed “ping”. This transducer is typically hull-mounted, while the pulse is emitted horizontally or vertically (toward the bottom of the water body) inside the water creating an acoustic beam, in a manner similar to a flashlight beam. When the sound wave hits (or encounters) a target, such as a fish, a part of the sound energy is reflected back to the receiver and interpreted as an echo. The time interval between the signal transmission and echo reception can be converted to the fish distance using a good approximation of the speed of sound in water, i.e. ~1500 m/s). Because the sound energy is diminished while traveling through the water, amplification is applied as a means of compensation for energy absorption and geometrical spreading.

In order to establish an accurate conversion, echosounders are calibrated. During this process, absolute echo levels are determined as averages of a number of transmissions. This technique uses echo integration and produces a value that is representative of fish density as a proportion expressed according to the principle of linearity (Foote, 1983). Fish densities calculated this way are averaged and interpolated along specific transects surveyed by vessels, in order to produce estimates of fish abundance with respect to the totally surveyed area (Simmonds et al., 1992). However, in order to obtain proper analysis results, the target acoustic properties have to be known in advance. As these vary depending on fish species, relationships are used between fish species and target-strength, which are established by identifying the ensonified fish and studying the acquired data with respect to the identified target characteristics. Echo characteristics are usually enough to match observed targets to fish species (Reid, 2000). However, determining additional information, such as age, length or the composition with respect to various other characteristics (such as maturity), often needs further confirmation. Acoustic surveys provide data for stock assessment, usually used to determine abundance classified by age. However, other classifications are also used for the same reason, such as spatial distribution, absolute biomass or average weight-age relations.

Acoustic techniques are very popular in fisheries around the world, with a multitude of sonars being available. Differences include wider beam angles and fully flexible beam orientation capability that enables the ensonifying of any directional aspect. Sonars can detect fish targets at relatively large distances, which renders them invaluable for studying fish behavior under a multitude of perspectives and aspects, not solely related to the original purpose of simply detecting the fish targets.

1.1.1. DIDSON System Characteristics

An acoustic camera DIDSON is a multibeam high-frequency sonar capable of producing high-resolution images through a unique adapted acoustic lens system. DIDSON, developed by the University of Washington, U.S.A. Applied Physics Lab can generate high-quality video sequences (Belcher et al., 2001, 2002a). DIDSON technology was developed in the context of harbor surveillance, in order to identify underwater intruders. It can be mounted on a submersible base or under a small vessel, which can be sent to produce a more detailed identification, upon coarse detection from a wider surveillance system. Among the original uses of the device were the assistance in diver supervision in conditions of high water turbidity (Elliott, 2005), the detection of mines or other obstacles (Belcher et al., 2002b), the monitoring of underwater constructions (Belcher, 2006) and the inspection of hulls from marine vessels (Vaganay et al., 2005). Modern sonar systems are used in the study of fish migration, with promising results (Maxwell and Gove, 2004; Baumgartner et al., 2006; Petreman et al. 2014). The high-frequency DIDSON operating mode utilizes 96 acoustic beams and is capable of mapping details such as the body shape, outline and fins of target fish

Acoustic cameras utilize very high frequencies, thus higher frequencies are reflected from softer parts of targets (muscle tissue, fins) and due to beam arrangement with single, narrow beams can provide image of the whole body of fish. As a result, fish behavior and morphology can be studied in much better levels of detail than what common sonar technologies can offer (Baumgartner et al., 2006). Additionally, noise of various types can also be mapped and recognized visually (e.g. bubbles, suspended matter, etc.). Due to the distinct shape of fish and their characteristic motions, which are typically non-linear, they can be distinguished from debris that typically exhibit a constant direction and velocity within adequately small time frames). In that respect, fish size measurements can be obtained directly from images of fish body shapes juxtaposed with the recorded data, without having to convert to or from target strength, which entails additional uncertainties and inaccuracies.

What is more important is the good performance of the system in the absence of light or in the presence of turbidity and, generally, in conditions of poor visibility. The DIDSON data are used for detecting, counting and measuring fish targets, greatly

facilitating operations such as the monitoring of fish behavior during, e.g., spawning, feeding, or migration.

The acoustic lens used has the capability to focus on targets located even at 1 m distance from the transmitter. Two emission frequencies are available, 1.8 MHz (the high-frequency mode) and 1.1 MHz (the low-frequency mode). In the high-frequency mode, the total beam is separated into 96 sub-beams (0.3° x 14°, horizontal x vertical dimensions), while the range is adjustable between 12 and 15 m (Holmes and Cronkite, 2006; Hughes, 2012). In the low-frequency mode, the beam is separated into 48 sub-beams (0.5° x 14°, horizontal x vertical dimensions) (Figure 1). The total ensonifiable field spans 29° x 14° (horizontal x vertical dimensions) in both modes (Maxwell and Smith, 2007). Long-range DIDSON cameras have the capability of increasing the range settings up to 33 m for their corresponding high-frequency mode (1.2 MHz) and up to 80 m for their corresponding low-frequency mode (0.7 MHz). The width of the pulse (duration between successive pulses) depends on the range setting and varies between 4.5 and 144 ms. As a result, frame rates up to 21 frames-per-second are supported. The accompanying data visualization application used for control and playback resembles a digital video program (Maxwell and Gove, 2004).

Data are visualized in 2D, resolved using X as the horizontal dimension and Y as the range dimension. The data display does not visualize the Z-dimension (Hughes, 2012). When the beam direction faces the bank (horizontal orientation), the DIDSON camera supplies relative distance and motion direction information regarding the detected fish but not their absolute positions with respect to the water column. The update rate is variable between 21 and 5 frames per second, according to the operating frequency and set range. The image is constructed throughout 4 cycles of successive transmission/reception, again in sets of 12 beams. In general, the low-frequency mode is utilized for longer ranges, over 12 m, but because half cycles are needed with respect to the high-frequency mode, the frame rates are similar between the two modes. Table 1 lists the DIDSON specifications.

DIDSON provides a remote focus function, which enables setting the focus on objects at a distance between 1 and 40 m from the sonar. When a range span is chosen for an image, DIDSON automatically sets the center of the set span as the optimum focus. This auto-focus can be overridden to any other range within an image. In all cases, the focus is technically good across the entirety of the settable range span.

Table 1. Detailed description of DIDSON specifications. (Source: Belcher et al. (2002b)).

Low Frequency mode	
Operating Frequency	1.0 MHz
Beamwidth (two-way)	0.6º horizontal by 14º vertical
Number of beams	48
Source Level (average)	205 dB re 1 µPa at 1 m
Range Settings	
Start range	0.75 m to 23.25 m in 0.75 m intervals
Window Length	4, 5, 9, 18, 36 m
Range bin separation	8, 17, 35, 70 mm respective to window length
High Frequency mode	
Operating Frequency	1.8 MHz
Beamwidth (two-way)	0.3º horizontal by 14º vertical
Number of beams	96
Source Level (average)	205 dB re 1 µPa at 1m
Range Settings	
Start range	0.4 m to 11.63 m in 0.4 m increments
Window Length	1.1, 2.2, 4.5, 9 m
Range bin separation	2.2, 4.4, 9, 18 mm respective to window length
Both Modes	
Frame rate	4-21 frames
Field of view	29º
Remote focus	1 m to maximum range
Power consumption	30 watts (using 115VAC or 14-18VDC)
Depth rating	152 m (500 ft) or 2400 m (8000 ft)
Control	Ethernet, RS232 or switches
Output format	Ethernet and NTSC video
Cable length	152 m
Dimensions	30.7 cm long by 20.6 cm high by 17.1 cm wide
Weight in air	7 kg
Weight in water	0.8 kg
Topside requirements	Computer running windows, Ethernet card and video monitor

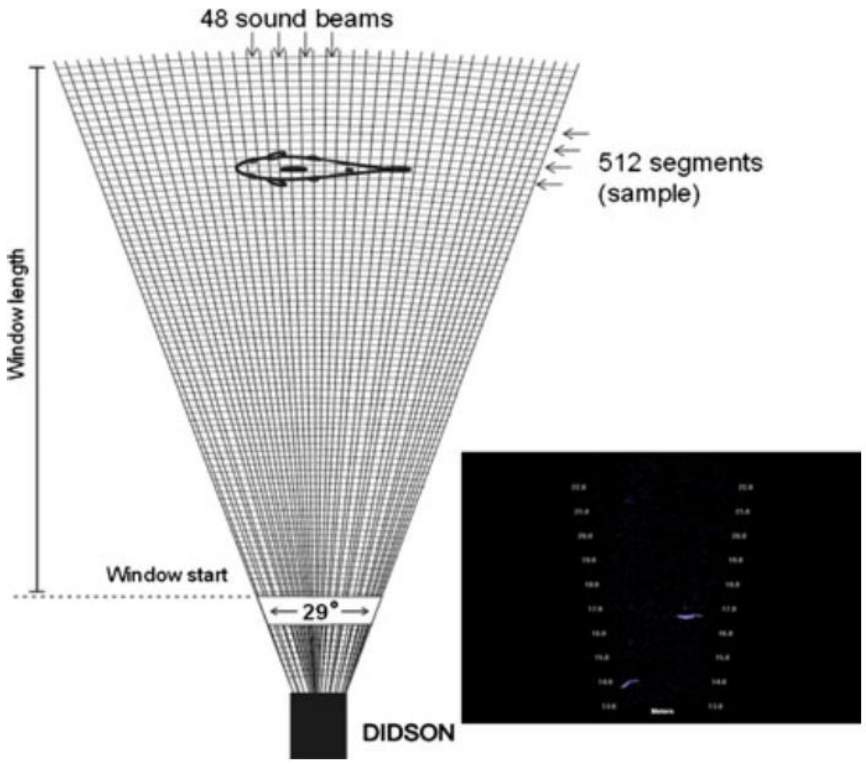

Figure 1. A brief description of the formation procedure of a DIDSON frame, operating at low frequency (Source: Lilja and Orell, 2011; Martignac et al., 2015).

1.2. Review of the data processing methods

The most widespread applications of DIDSON in the field of fisheries research concern the study of fish counting (e.g. Han et al., 2009; Faulkner et al., 2009; Pipal et al., 2010), fish sizing (e.g. Han et al., 2009; Burwen et al., 2010; Becker et al., 2011), fish behavior (e.g. Mueller et al., 2006; Johnson et al., 2012; Bevelhimer et al., 2015) and monitoring fish populations (e.g. Galbreath and Barber 2005; Maxwell and Gove, 2007; Hughes, 2012; Martignac et al., 2015;).

However, freshwater ecosystems are very diverse and environmental factors vary among hydroacoustic recordings. Therefore, efficient analysis software is required to handle and process DIDSON data at high speed. To date, several analysis software applications have been developed, which are capable of enumerating and sizing fish and investigating fish behavior from DIDSON data. Namely, DIDSON software (Sound Metrics Corp.), ECHOVIEW (Myriax Pty Ltd, Hobart, TAS, Australia) and Sonar5-Pro

(Lindem Data Acquisition, Oslo, Norway) have been used among diverse groups of scientists for various objectives such as fish stock assessment, fish monitoring, and environmental management. Sonar5-Pro (CageEye, Oslo, Norway) offers tools to display echograms, track fish echoes automatically and extract tracks to a database. Processing tools include background subtraction, automatic tracing and classification or single echo detection. Manual tracking, track editing and fish length ruling is also available (Balk and Lindem, 2002; Balk et al., 2009). Like the Sound Metrics Corp. software, Sonar5-Pro makes DIDSON data (echogram and video displays) visible, manually tracking fish echoes and extracting information into a database. An automatic tracking tool is available. It has several settings (e.g. number of consecutive pings, number of ping gaps, cluster size) and enables consecutive echoes to be grouped into a track. Sonar5-Pro can extract all the available data in DIDSON files. Extraction of fish data from DIDSON records. Similarly, Echoview software also visualizes and exploits DIDSON data. Some Echoview tools perform semi-automated data analysis (Han et al. 2009; Kang 2011) and some are similar to those of DIDSON software, such as the background subtraction tool. All these programs count fish passages, provide estimates of fish length (Han et al. 2009) and can describe fish behaviour (Kang 2011).

As the DIDSON data can be properly converted to an image-snapshot stream similar to a video, most modern methodologies employ computer vision techniques (e.g. Langkau et al.2012) to resolve motion patterns from image sequences. While having substantial success (e.g. Mueller et al, 2008; Han et al., 2009), these methods almost invariably necessitate human intervention, in order to refine the final results by removing artifacts or other random effects. Furthermore, while automation is practically attained, procedures are typically based on extensive parameterization and it takes significant expertise and experience to suitably fine-tune all processing parameters for an optimal result. Finally, because underlying algorithms are complicated and work based on a varying number of assumptions, their performance is almost invariably situation-specific. Fuzziness in the data, minimum and maximum expected target size, average single-target shape and average target separation distance are, among others, just a few of the variables that can affect the capability of algorithms to detect and track single targets throughout a dataset.

1.3. Objectives

The primary aim of this thesis is to explore and improve upon current methodologies for automated fish target detection and tracking from DIDSON data. The first objective is to propose and test the efficiency of a relatively recently established computer vision technique that has been considered and used in various suggested or experimental motion detection scenarios such as (Wu, 2012; Alqaddafi, 2019) that of **optical flow**, for the purpose of improving the detection and tracking of motion in DIDSON data. This will be achieved by establishing and describing a specific workflow with optical flow calculations integrated into it, identifying all parameters relevant to the quality of the outcome and highlighting their importance and interactions with other characteristics of the involved dataset, instruments or other circumstances. The second objective is to propose and test a methodology to automate the process of parameter selection for a subset of the involved parameters in the employed workflow. This will be achieved by coupling the workflow with a parameter re-adjustment method, specifically a genetic algorithm. The proposed framework will be applied as a case study in a DIDSON dataset which was obtained from an acoustic monitoring recording in the Vltava River.

2. Study site and Data collection

The fish migration data for this study were obtained from a stationary acoustic recording conducted in the Vltava River, Czech Republic, under the project "Coexistence of human and pearl mussel Margaritifera margaritifera in the Vltava River". A DIDSON was deployed at one site on the Vltava River to monitor fish passage from the Lipno reservoir to the Vltava River, in the area of the Šumava National Park.

In order to estimate the migration timing and abundance of fish, the fish passage monitoring was performed for two spawning periods (5/4/2014-9/5/2014 and 18/3/2015-18/6/2015). During these periods the acoustic recording was almost continuous. In the cases of high flooding risk, especially at the end of March and April, the acoustic camera was temporarily removed for a few days, for safety reasons. Once the fish upward movement was no longer noticeable in the recordings and the fish migration of most species had ceased, the continuous monitoring was terminated. Water temperature and flow velocity were additionally measured.

The DIDSON unit consisted of the sonar, a set-top control box, a data cable, control software and an associated laptop computer. The DIDSON was directly connected to the set-top box, which was linked to the laptop via an Ethernet connection. The image was transferred from the unit to the laptop via the control software, which displayed the data as a streaming image.

The DIDSON acoustic camera was operated in the high-frequency mode. Operation with the high frequency provided an optimal footage. The DIDSON recordings have sufficient resolution to identify of different objects and direction of movement, but not necessarily species unless there are clear and consistent differences in body features or behavior.

In order to obtain the highest quality image, care had to be taken when selecting the monitoring sites and mounting the DIDSON acoustic camera. Structure of the bottom, passing debris or other structures creating shadows and therefore 'blind zones' had to be taken into consideration. A cross profile of the river with the evenly increasing depth of the river bank close to the cabin area of Nová Pec, about two kilometers above the river mouth (48°48.52115′ N, 13°56.77817′ E) was selected for the mounting of the acoustic camera (Figure 2). At the selected site, the slope of the

right river bank is steeper than the slope of the left river bank. In addition, the gravel river bottom is finely shaped to a relatively smooth surface, without major obstructions. A platform for holding DIDSON was placed approximately three meters from the shore of the right riverbank. DIDSON had a cross-sectional orientation with respect to the river current to the opposite side of the river, and its lower edge of the beam horizontally followed the bottom from the shallowest and deepest part of the river. Two guiding fences were used to guide migrating fish away from the shore, where their detection by an acoustic camera would be difficult, into a limited space observed by DIDSON (Figure 3, 4). In addition, a small fence (30-40 cm high) was placed along the bottom, between the two guiding fences, in order to prevent fish from passing over just above the bottom.

Study site

Legend

- Study site

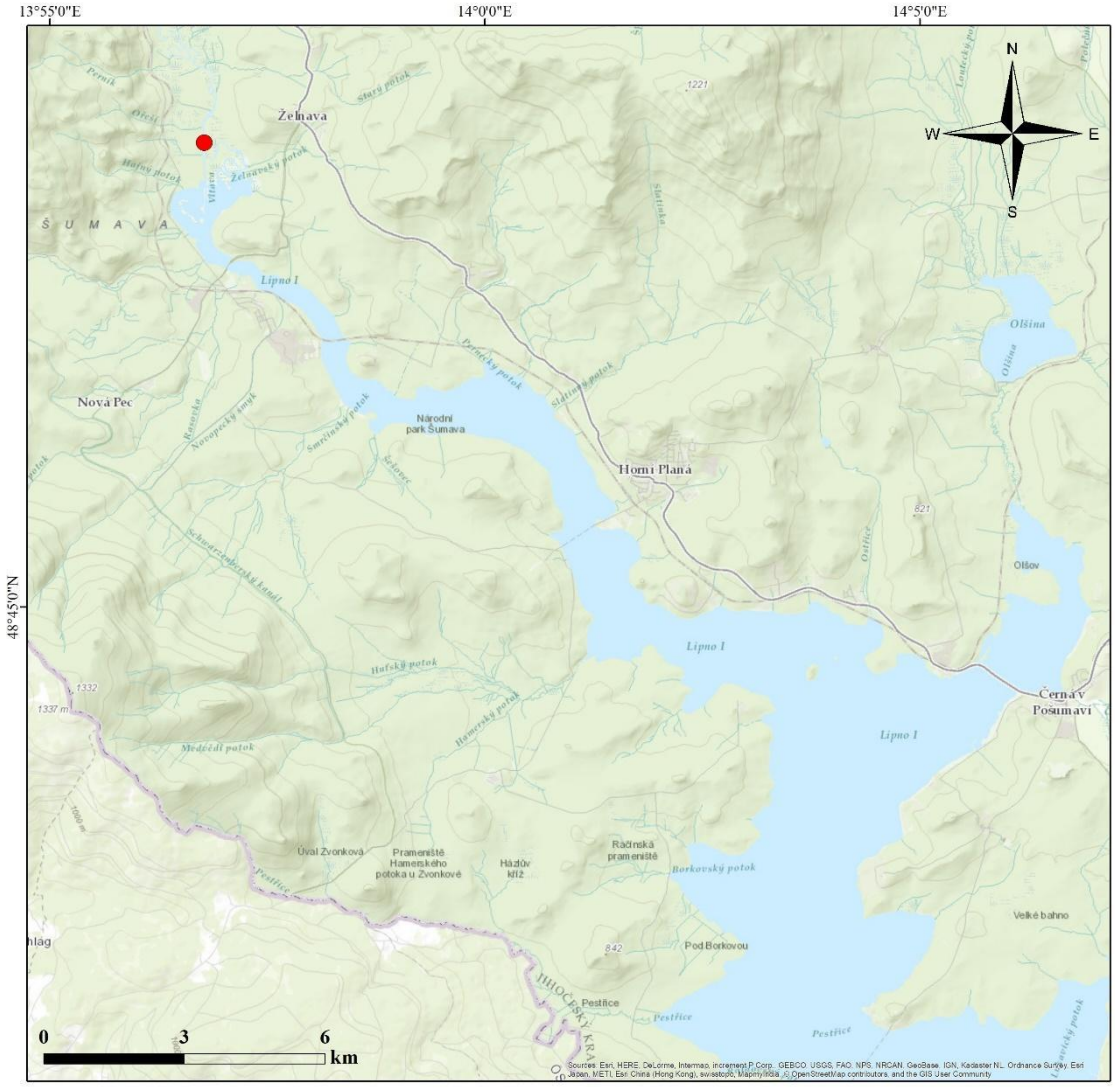

Figure 2. Location of the monitoring station (solid dots) in Vltava River, for two spawning periods (5/4/2014-9/5/2014 and 18/3/2015-18/6/2015).

Figure 3. The arrangement of the guiding fences during the first spawning period (5/4/2014-9/5/2014) (Source: No. P14V00000374 Soužití člověka a perlorodky říční ve Vltavském luhu, Operational Programme Environment).

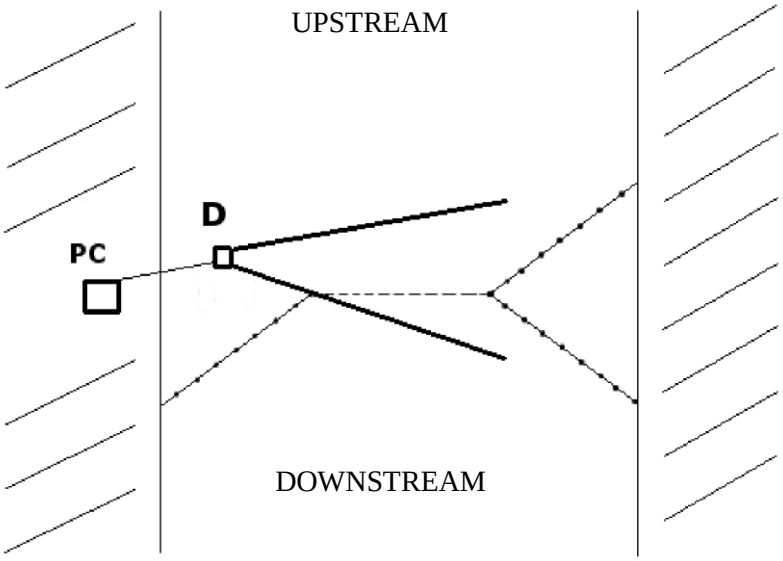

Figure 4. Diagram of the equipment set-up (PC – placement of the operation unit, D – DIDSON with an observable space, -- bottom fence,–– guiding fences, D: DIDSON) (Source: No. P14V00000374 Soužití člověka a perlorodky říční ve Vltavském luhu, Operational Programme Environment) .

3. Methodology

3.1. Theoretical Background

3.1.1. Optical Flow, Motion Detection and Problem Statement

Horn and Schunck (1981) define optical flow as the "*distribution of apparent velocities of movement of brightness patterns in an image*". Considering point brightness as the main component of visual representation, optical flow can be described as the brightness velocity field. Given two images at instants t and $t + \Delta t$, the brightness velocity field at each point represents motion and is an important measure to detect the magnitude and directions of apparent motion.

Considering the typical visual representation of an image as an array of picture elements (pixels) with specified brightness (intensity) values $B(x,y)$ and given that the image is a representation of coherent and temporally sequential states, the effects of motion are represented as the introduction of the time variable in the brightness field of the image, leading to $B(x,y,t)$. This brightness field is usually the available data and is represented as an image sequence. Considering that this sequence is the result of observation of a specific scene under a given set of conditions, the brightness field directly represents the composition of objects that make up the scene. The goal of optical flow methods is, then, to extract the potential displacements of objects on the scene based on the visual representation of the temporally variable brightness field.

To make the determination of motion possible from an image sequence, a few fundamental assumptions are made [Horn and Schunck, 1981]. The first and most important assumption is that the effect of physical motion is translated as a corresponding translation of a pixel intensity to a different, typically neighboring, pixel:

$$H(x, y, t) = H(x + \Delta x, y + \Delta y, t + \Delta t) \quad \textbf{(1)}$$

This is called the *brightness constancy constraint* (Horn and Schunck, 1981; Fleet and Weiss, 2005) and practically entails the assumption that the surface radiance of all objects forming the represented visual scene is constant. Although this is untrue for real-world scenes, this assumption is remarkably adequate for most applications (Fleet and Weiss, 2005). Because there is only one intensity value at the point (x,y) and it is theoretically impossible to determine two unknowns (Δx, Δy) from a single measurement

(*Δt* is known as the image sequence interval), various methods make different additional assumptions in the form of constraints, in order to bypass the indeterminacy and arrive to a solution. Considering the Taylor expansion of (**1**) around (*x*,*y*):

$$H(x,y,t) = H(x,y,t) + \frac{\partial H}{\partial x}\Delta x + \frac{\partial H}{\partial y}\Delta y + \frac{\partial H}{\partial t}\Delta t + \varepsilon, \qquad (\mathbf{2})$$

where ε represents terms of order higher than 2, and dividing by Δt, ignoring higher order terms, and taking the limit as $t \to 0$, the *gradient constraint equation* or *optical flow constraint equation* is retrieved (Horn and Schunck, 1981; Beauchemin and Barron, 1995; Fleet and Weiss, 2005):

$$\frac{\partial H}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial H}{\partial y}\frac{\partial y}{\partial y} + \frac{\partial H}{\partial t} = \nabla(H)\cdot\vec{u} + H_t = 0, \ (\mathbf{3})$$

where

$$\vec{u} = (u_x, u_y) = \left(\frac{\partial x}{\partial t}, \frac{\partial y}{\partial t}\right), \ (\mathbf{4})$$

is the velocity vector. Equation (**3**) more clearly exhibits the aforementioned indeterminacy, namely that it is impossible to determine u_x *and* u_y at the same time, from a single brightness value *H* at a specific pixel, as this is one equation with two unknowns. This algebraic manifestation is called the *aperture problem* and arises naturally in the study of motion on a visual representation when only local structure is available (Ullman, 1979; Bertero et al., 1988; Beauchemin and Barron, 1995).

3.1.1.1. The Aperture Problem – Ill-posedness of Optical Flow Computations

The aperture problem is, mathematically, based on the interpretation of equation (**3**) and means that brightness values and their changes can only lead to calculation of the velocity component that is *normal* to the brightness pattern, i.e. *parallel* to the local brightness gradient direction:

$$\nabla(H)\cdot\vec{u}=\|\vec{u}_{\perp}\|=-H_t \Leftrightarrow \vec{u}_{\perp}=-H_t\frac{\nabla(H)}{\|\nabla(H)\|^2}. \quad (5)$$

The physical interpretation of the aperture problem stems from the fact that calculations and visual interpretation typically take place in reference to small local visual neighborhoods similar to looking the entire image through an aperture (hence its name). It follows that if not enough local structure visualization is available (mainly due to ignoring the distant surrounding structures), as is the case, for example, for textures aligned with a single direction/orientation, such as stripes, it is impossible to determine velocity components in other directions (Beauchemin and Barron, 1995), see e.g. Figures 5 and 6.

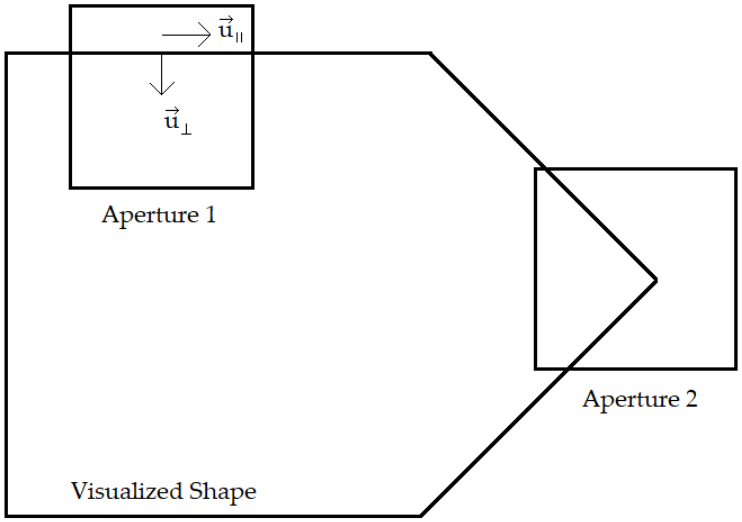

Figure 5. Example of a visualized shape as seen through apertures at 2 different locations.

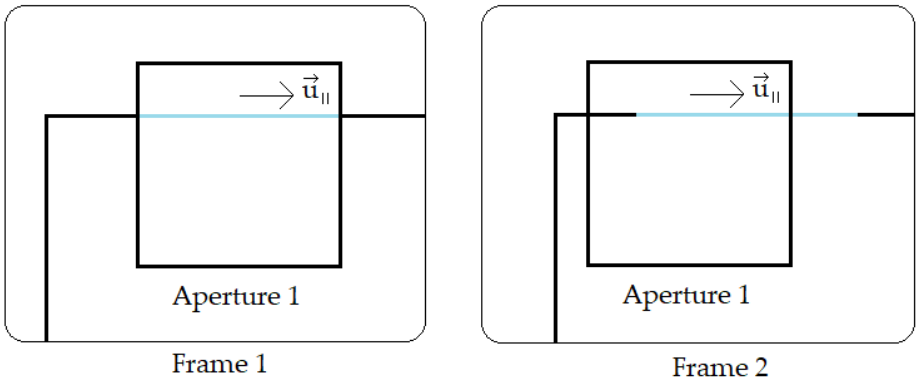

Figure 6. Two consecutive frames of the moving shape of Figure 1, considering Aperture 1 as a steady location on the visualization plane (video). The actual segment displacement between frames is painted with light blue.

In Figure 6, the motion of the highlighted segment cannot be determined when only the viewport of Aperture 1 is observed because the end result is two identical visualizations. As a result, it is impossible to quantify the total horizontal displacement of the shape by observing only through Aperture 1, because brightness does not appear to change. Motion in the direction that is perpendicular to the texture (shape) orientation, i.e. in the brightness gradient direction, however, can be readily determined because the displacement would be discernible and measurable between the different brightness patterns in that direction, as seen in Figure 7.

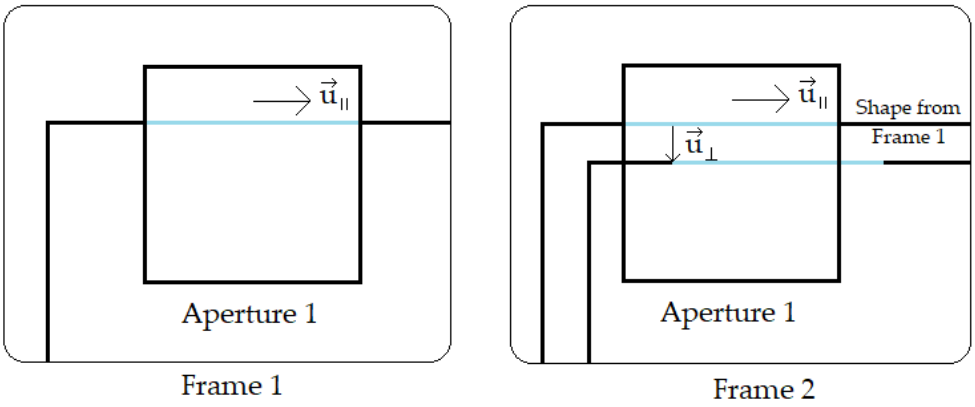

Figure 7. Two consecutive frames of the moving shape of Figure 1, together with Aperture 1. Motion in the perpendicular direction of constant brightness is directly quantifiable.

In Figure 7, the motion in the direction of the brightness gradient is measurable (considering the unit of time between the two frames for simplicity), but it still remains unclear how much the pattern has moved in the horizontal direction because it is impossible to match pixels, as all pixels have (normally almost) identical brightness values. The aperture problem can be resolved if adequate local structure is available through an Aperture, such as Aperture 2 from Figure 5.

In Figure 8, it is clearly shown how more elaborate brightness patterns having a more complete local structure, expanding in more than one direction, can lead to resolution of the aperture problem. However, this depends on the case and when determining velocity of structures, local brightness pattern structure is not evident *a priori*, therefore the aperture size is relevant.

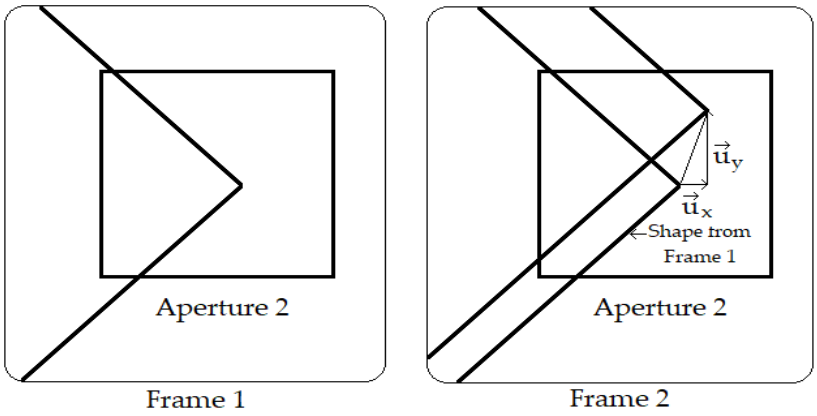

Figure 8. . Two consecutive frames of the moving shape of Figure 1 together with Aperture 2. Because local structure visible through Aperture 2 is more elaborate, local shape displacement can be more accurately calculated between the two frames.

The illustrations of Figures 5, 6, 7 and 8 imply the additional assumption that the shape translation represents a corresponding movement of its real-world object counterpart, which is purely parallel to the image plane. However, it is not infrequent for motion on real-life captured video sequences to represent actual movement with a nonzero component perpendicular to the image plane, i.e. depthwise with respect to the visual sensor. This, along with the aforementioned assumptions, are taken to hold "locally" on any visual representation and the degree of accuracy of optical flow estimation is, therefore, directly linked to how far a processed image sequence deviates from these assumptions (Beauchemin and Barron, 1995).

The aperture problem and the inability to determine actual motion when local structure (as brightness patterns) degenerates in terms of dimensionality, is what makes the optical flow computation an ill-posed problem according to the definition of Hadamard (Hadamard, 1902). Starting from a specific frame and considering a local neighborhood (aperture), many different velocity fields could lead to the same "next" frame when structure degenerates. This way, the inverse problem does not have a unique solution (Bertero et al., 1988; Bruhn et al., 2005).

3.1.1.2. Optical Flow Algorithms

A number of differential methods have been developed that introduce additional constraints, in order to make the optical flow calculation feasible. For example Horn and Schunck assume smoothness in the velocity field by considering that all motion is either

rigid, or deformation-based and discontinuities mainly occur at occlusion points, i.e. edges between distinct object borders on the image (Horn and Schunck, 1981). This *smoothness constraint*, is a general assumption that enables the gradient-based estimation of optical flow.

An important distinction in the literature is made between local and global methods. The difference lies in attempting to minimize local or global expressions that are used as energy functions (or functionals). While local methods do not result in dense optical flow fields, they perform better under noise, whereas the opposite is true for global methods, i.e. they produce dense and continuous flow fields but have higher sensitivity to noise. The most popular representatives of algorithms in the two categories are the Lucas-Kanade local method (Lucas and Kanade, 1981; Lucas, 1984) and the Horn and Schunck global method (Horn and Schunck, 1981), both of which have various re-adaptations in the literature, while combinations of local and global methods have also been proposed (Bruhn et al., 2005).

One of the more recent optical flow calculation algorithms is based on Gunnar Farnebäck's motion estimation techniques using quadratic polynomials. In specific, using frame pairs, every neighborhood of each image in the pair is approximated using quadratic polynomials. The displacement field is, then, calculated by employing suitable adjustments of a general translation-based polynomial transformation model (Farnebäck, 2003). Farnebäck attempted to compensate for background motion, in order to obtain reliable and smooth motion fields for sequential frame pairs acquired, for example, under high-frequency vibrations of the capturing medium. The novelty of Farnebäck's motion estimation algorithms was the use of orientation tensors, as well as the inclusion of parametric motion models. Calculations carried out in the frame of this thesis are based on the MATLAB® scripting development environment's implementation of the Farnebäck optical flow estimation algorithm.

3.1.2. Genetic Algorithms – Evolutionary Methods

Genetic algorithms are a mathematical abstraction, which can be seen as the algorithmic counterpart to evolution. A problem modelled as a parameterized system can be said to have, as solutions, various parameter sets, which may adhere to various constraints. Genetic algorithms are, then, a class of methods that aims to find better

approximations to such solution sets, given a number of initial sets, along with an evaluation criterion. Evaluation criteria generally fall in two broad categories, namely *fitness* or *penalty* type criteria. Evaluation of solutions through fitness express a problem of maximization, that is, the search for the solution set with as large fitness as possible. Similarly, evaluation of solutions through penalty is analogous to a minimization problem, i.e. the search for a solution set with the minimum possible penalty with respect to all other solutions. The two categories of criteria are practically equivalent in that each type can be adapted to the other one through suitable mathematical transformations (e.g. the inverse of penalty can be taken as the fitness).

An important note is that the aforementioned solutions are not necessarily the optimal solutions to the corresponding problems. In fact, because genetic algorithms mimic evolutionary processes, it is not always possible to consistently reason about them at a mathematical level. From a simplified perspective, a genetic algorithm seems to resemble a black box, which tends to consistently produce improved solutions to a given problem, based on a specified evaluation criterion. Beside their cryptic nature, a number of theorems and analyses have been carried out with respect to genetic algorithms. As a major example, the *Schema Theorem* (Holland, 1975; Lankhorst, 1996) is a first proof that solutions can consistently and favorably (from a computational perspective) converge towards increasingly improved solutions, as long as existing solutions are better than average. Holland (1975) also formulated what is known as the *canonical genetic algorithm*, which uses *binary strings* to encode solution parameters into an *artificial chromosome*, which is the informational counterpart of a biological chromosome.

One of the necessary prerequisites, in order to be able to solve a problem using a genetic algorithm, is the *feasibility* of evaluating solutions, regardless of the feasibility of the problem itself. The breakdown of a typical genetic algorithm results in the following steps:

- Expressing the solution set parameters in terms of *chromosomes* (i.e. *encoding the solution parameters*).
- Deriving a set of initial solutions (even random).
- Applying a series of *genetic operators* (such as *crossover* and *mutation*) and recombining solutions.

- Proceeding to the next generation (which is modelled as an iterative process) by applying a *selection* strategy (e.g. *elitism*, random, based on fitness-proportions, or a combination thereof).

Followed in sequence, these steps closely resemble biological evolution and lead to the continuous improvement of solution sets as generations pass. This can, therefore, be seen as an optimization problem, in which a swarm of solutions iteratively approaches an optimal region on the solution landscape of the problem. A concern, however, was that the convergence to an optimal solution, rather than "wandering" around a wider optimal region, was not always guaranteed, as only the canonical genetic algorithm had been analyzed and attained a relatively stronger mathematical foundation (De Jong, 1993). As a result, a number of modifications appeared to be warranted, in order to efficiently approach optimality, while not wandering off of it. As an example, Rudolph (1994) indicates that, in order to guarantee convergence, it is necessary to employ *elitism*, i.e. the retaining of the best solution from a given generation to the next. Other modifications of genetic algorithms have been proposed and applied with varying degrees of success, such as a combination with simulated annealing methods (e.g. Lozano et al., 1999).

An important application of genetic algorithms is their use in solving nonlinear mixed integer constrained-space problems (e.g. Li and Gen, 1996). As genetic algorithms do not require additional information regarding the problem domain or the behavior of the problem in various regions of the solution landscape, they are well-suited for hard problems with vague definitions and/or incomplete specifications. Integer problems are problems, where the solution set is potentially bounded and a subset of the parameters (or all of them) vary over the integers. Because some of the parameters related to the determination of optical flow are by definition integral values (e.g. the filter size and the neighborhood size), genetic algorithms are a good fit for determining locally optimal solutions for them in the context of a constrained mixed integer optimization problem.

3.2. Optimal Target Mask Extraction Algorithm

In order to obtain the desired information from the DIDSON data, a step-wise algorithmic process is performed. The steps of this process are described below. A workflow diagram that schematically represents the algorithm is also depicted in Figure 9. The algorithm is comprised of a standard part and an iterative part, which is based on a predefined iteration count *M* that is manually decided. In practice, because of the computational effort involved in the iterative part of the algorithm, this number is kept to a relatively small value. Additionally, part of the iterative process is based on choosing a number *N* of parameter sets, which are re-adjusted in the course of the algorithm, used for the calculations of *N* corresponding optical flow fields. This number also has a great impact on the computational effort involved in the algorithm as it represents the number of optical flow fields that have to be calculated at each iteration. As a result, this number is also kept to a relatively small value.

STANDARD PART OF THE ALGORITHM:

- Data extraction, which converts the raw DIDSON data to a usable image sequence that can be analyzed using techniques from the field of computer vision, image processing etc.
- Data pre-processing, in order to limit the effects of noise and other artifacts while maintaining the main information pattern depicted on each frame.
- Foreground/background identification, in order to separate the stationary parts of each frame (*background*) from the "changing" areas.

ITERATIVE PART OF THE ALGORITHM:

- Initialization of *N* parameter sets for the optical flow calculations.
- Calculation of *N* optical flow fields for the entire foreground sequence, in order to determine the areas where actual change is most notable.
- Refined mask extraction from the calculated optical flow field.
- Final masked result assessment and re-calculation (based on various techniques) of *N* parameter sets (both old and new) for the optical flow calculations.
- Iteration of the above steps for a preset number of times.

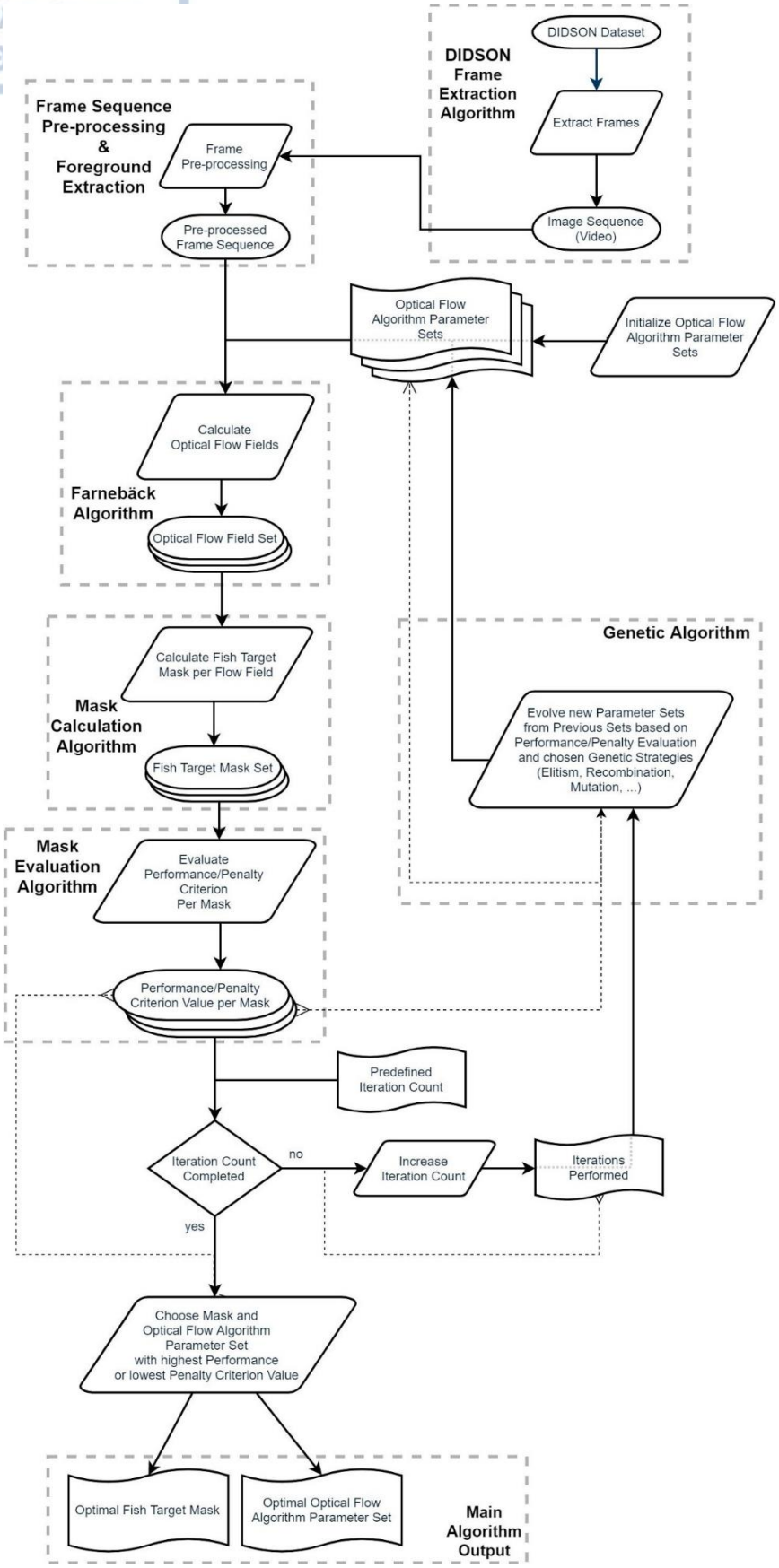

Figure 9. Flowchart depicting the proposed fish-target mask extraction from raw DIDSON data.

After the algorithm is completed, the output mask with the best assessment is the primary output, along with the optical flow calculation parameter set that led to that mask. This parameter set can be further tested for performance as a proxy parameter set for image sequences with similar characteristics. The following sub-chapters analyze each of the steps separately, presenting intermediate results of the calculations along the way to illustrate the various outputs of the algorithmic process. Apart from that, a specific identification of the parameters that can significantly influence the result is made, highlighting their importance, interactions with other dataset characteristics, as well as rationales for their determination.

3.2.1. Data Extraction

The raw DIDSON data are contained in corresponding files as successive data tables (representing pings), each one consisting of a beam-range indexed array of data points. Each element in this array corresponds to specific beam and range indices and is an 8-bit number, which usually represents an encoded version of the observed signal. The value corresponds to a specific location in space, which can be determined based on the data point's relative position to the instrument (taking the orientation of the instrument into account) and the precise external position of the instrument, which is usually acquired through a GPS sensor. The extraction of the data aims to transform it to a frame sequence similar to a video capture, which can then be used to further analyze apparent motion and detect targets by employing pertinent techniques from the field of computer vision. The characteristics of the sample used in this work are shown in Table 2 below.

Table 2. DIDSON sample data characteristics and parameters.

Parameter	Value
DIDSON Version	DDF_03
Frame Rate (frames per second)	8 fps
Recorded Range (m)	1.257 – 11.257
Total Recorded Range (m)	10
No. of Beams	96
Samples per Channel	512

In order to transform the data into a visual frame sequence, it is necessary to restore the *relative dimensional integrity* of the captured space, as well as the *relative spectral range* for each frame. Restoring the dimensional integrity will ensure that the visuals depicted on each frame are properly scaled representations of the real ensonified space, maintaining the correct aspect ratio in terms of their dimensions. Restoring the relative spectral range, on the other hand, will ensure that each frame is a proper representation of the reflected signal intensity at each volumetric element of the ensonified space. The dimensional correspondence between the ensonified space and the visual frame representation is based on using a pixel for each volumetric element of the ensonified space, whereas the spectral correspondence is based on using an intensity value at each pixel for each reflected signal intensity value at the corresponding volumetric element (*voxel*).

The data extraction pipeline includes reading the raw data files, appropriately converting the data samples to dB, building frame-arrays from the data samples and reconstructing images from the frames. Figure 10 below depicts a frame and the corresponding reconstructed image, while Figure 11 provides a schematic representation of the process. The reconstructed image can also optionally be flipped vertically to potentially resemble the visualization of the software accompanying the instrument.

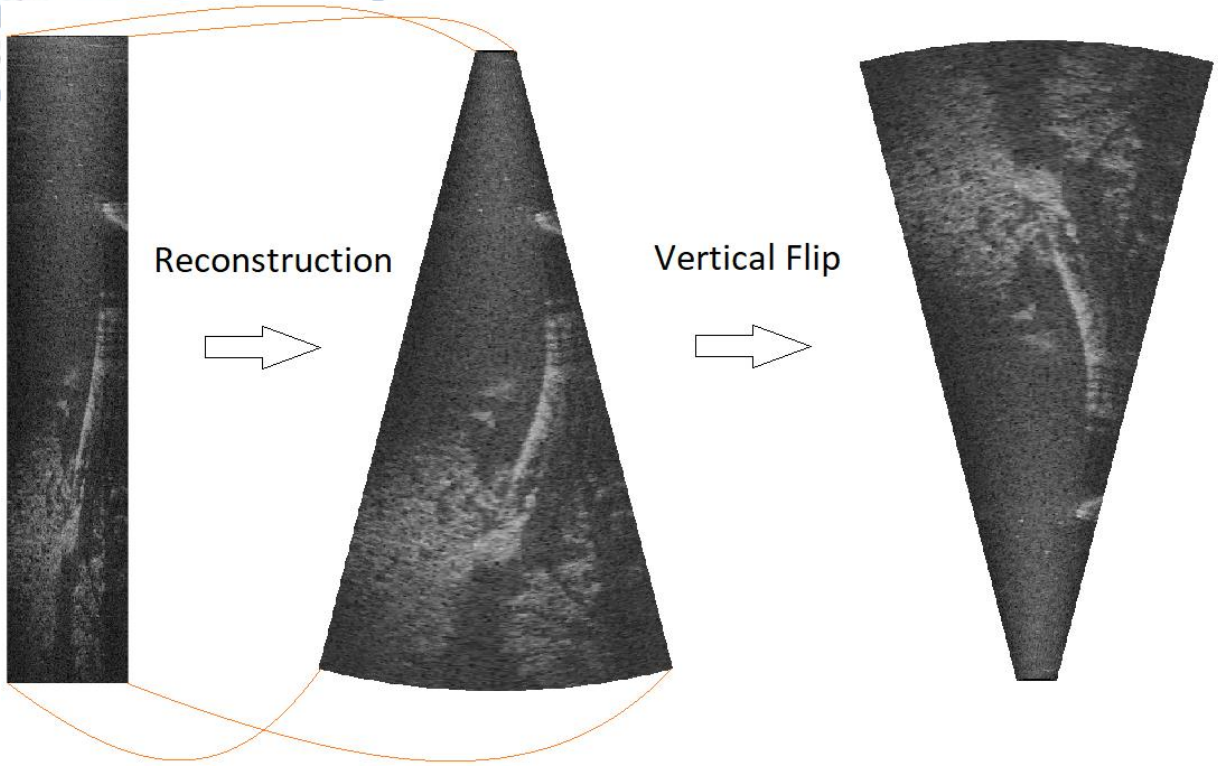

Figure 10. Reconstruction of frame to image (vertical flip to make image resemble the original view of the software accompanying the instrument).

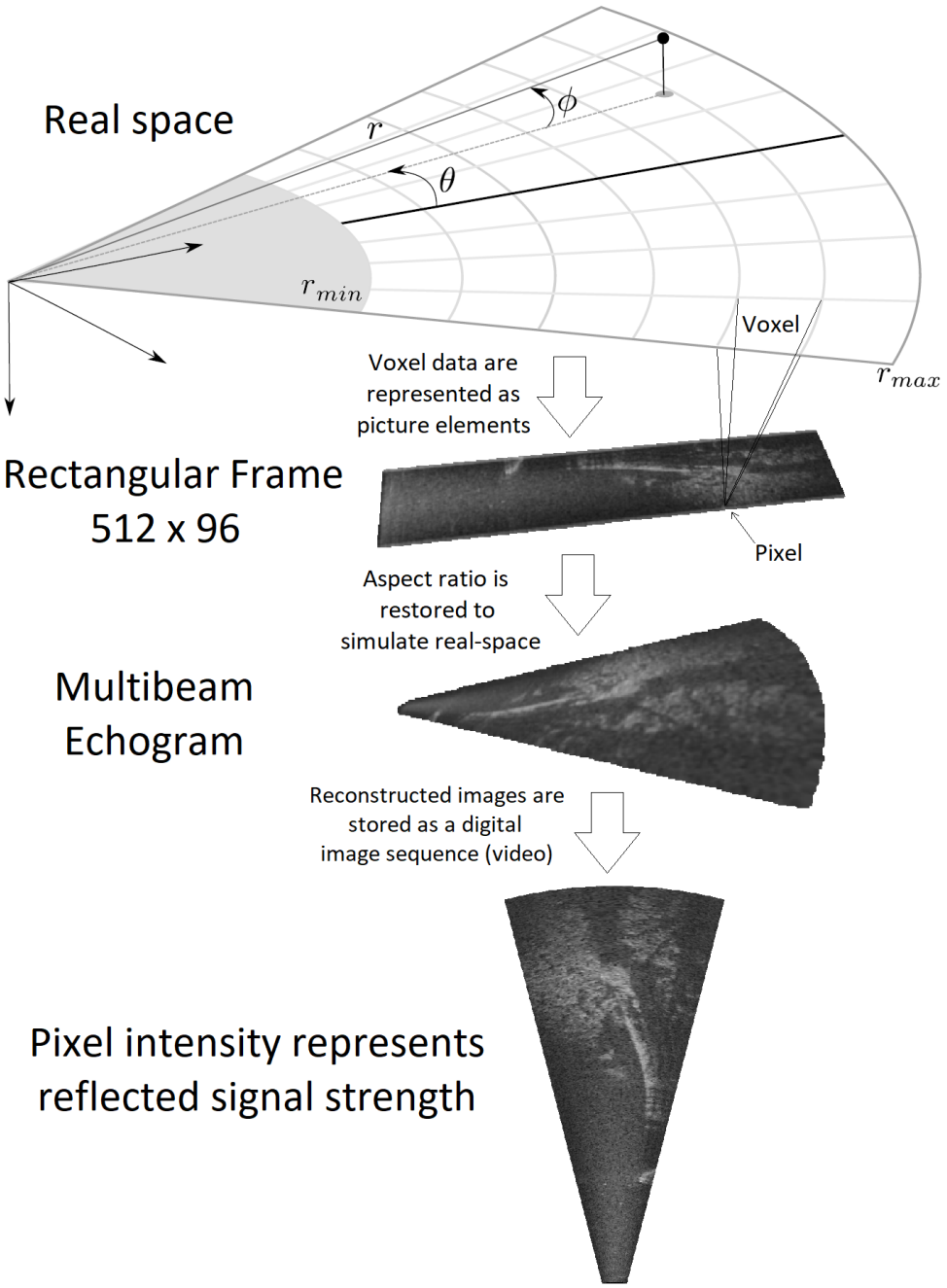

Figure 11. Conversion diagram of real-world ensonified Field-of-View to a digital image sequence representation (adapted, in part, from [Johannsson et al., 2010]).

Because the recorded data sample values for each voxel are *encoded*, i.e. converted to binary format suitable for storage in digital media (for example in 8-bit values, which represent integers from 0 to 255 for the **DDF_03** format), they can be converted and rescaled, in order to represent the true reflected signal power (dB). In the data extraction process, the encoded values are converted to Sv (MacLennan et al., 2002), based on the instrument specifications. DIDSON echosounders do not apply a time-varied gain to the

recorded signals, so, in the frame of this work, the data values are converted to Sv (volume backscattering coefficient) by applying a simple range-correction term to the recorded value as:

$$S_v = V_r + 20log_{10}(r), \quad \textbf{(6)}$$

where r represents the range of the data sample and V_r represents the recorded data value. Based on the total range of 10m (from Table 2 above), this can be expressed as 10 * (i / 512):

$$S_v(i,j) = V_r(i,j) + 20log_{10}\left(10\frac{i}{512}\right), \quad \textbf{(7)}$$

where (i, j) represent the row and column indices in a data sample array of a single frame (ping). Figure 12 shows the typical histogram of a frame from the sequence used in this work. The total pixel count of a frame is equal to 512 * 96 = 49,152. The histogram is derived from the frames rather than the images (leftmost depiction in Figure 10), because reconstructed images are based on interpolating and using virtual beams, which would skew the original distribution. Additionally, because of the necessity that all represented images are rectangular, additional white or black padding pixels are added outside the field-of-view, which would also create a very large count in the corresponding bins for value 0 or 255.

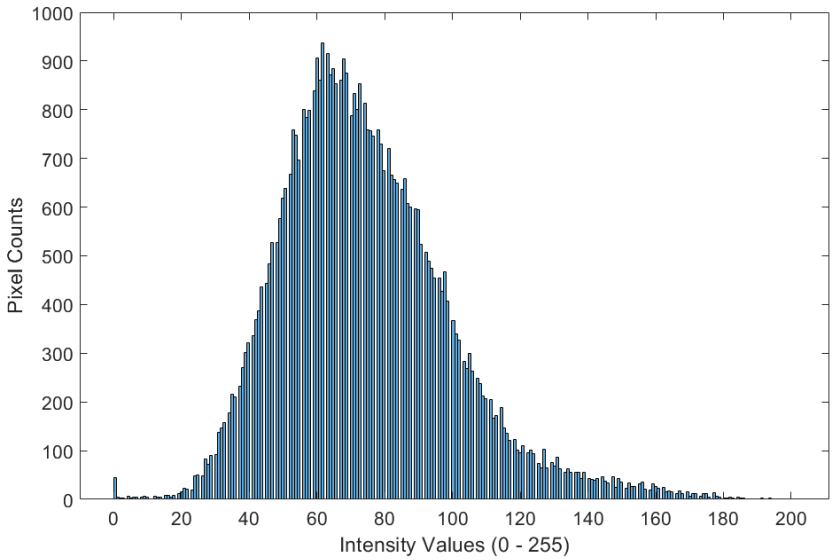

Figure 12. Typical histogram of an extracted data frame.

3.2.2. Pre-processing of Reconstructed Frame Sequence

After extraction of the raw data and conversion to the main frame sequence, standard computer vision techniques can be applied. A very important pre-processing step is the smoothing of the data, in order to minimize the effects of noise, which have the potential lead to increased outliers and mis-detections in the subsequent mask extraction process. Various methods are used in the literature, which can be applied spatially (within the same frame) or temporally (along the temporal dimension), or in combinations thereof (e.g. Bulas-Cruz et al., 1993; Simonoff, 1996). In the present work, a Gaussian temporal smoothing model was used, with a *frame–window* equal to1 second of the frame rate, i.e. *8 frames*. This parameter was determined after testing and the primary affecting dataset characteristic was determined to be the *fish target velocity* distribution throughout the recorded data.

Choosing a value for the temporal smoothing window is a trade-off between less noise and clearer, more easily identifiable targets. Therefore, it is an attempt to achieve an as pronounced smoothing and reduced noise effect as possible (for higher window values), at the expense of reduced contrast of faster target motions, thus making them appear as pale, “ghastly” imprints over the frame. At its heart, the Gaussian temporal smoothing algorithm based on a frame-window acts as a Gaussian-weighted moving

average that uses the frame-sequence contained in the window. In the present work, the algorithm was applied by extracting the time series $H_{i_0,j_0}(t)$ of each pixel as the array values in the temporal (third) dimension as:

$$H_{i_0,j_0}(t) = H(i,j,t)\big|_{i=i_0,j=j_0}, \quad \textbf{(8)}$$

applying the Gaussian moving-average smoothing method to get the smoothed time series $\hat{H}_{i_0,j_0}(t)$ and then reconstructing the original image sequence by suitably arranging the smoothed 1D time series as the third dimension. In effect, the results are presented in Figures 13 and 14, below, which depict the same reconstructed DIDSON image frame before and after temporal smoothing using a window of 12 and 8 frames, respectively. To better illustrate the effects of smoothing, the frame of a reconstructed DIDSON image with an identified fish target on it has been deliberately chosen, with the target highlighted with a yellow circle on all images.

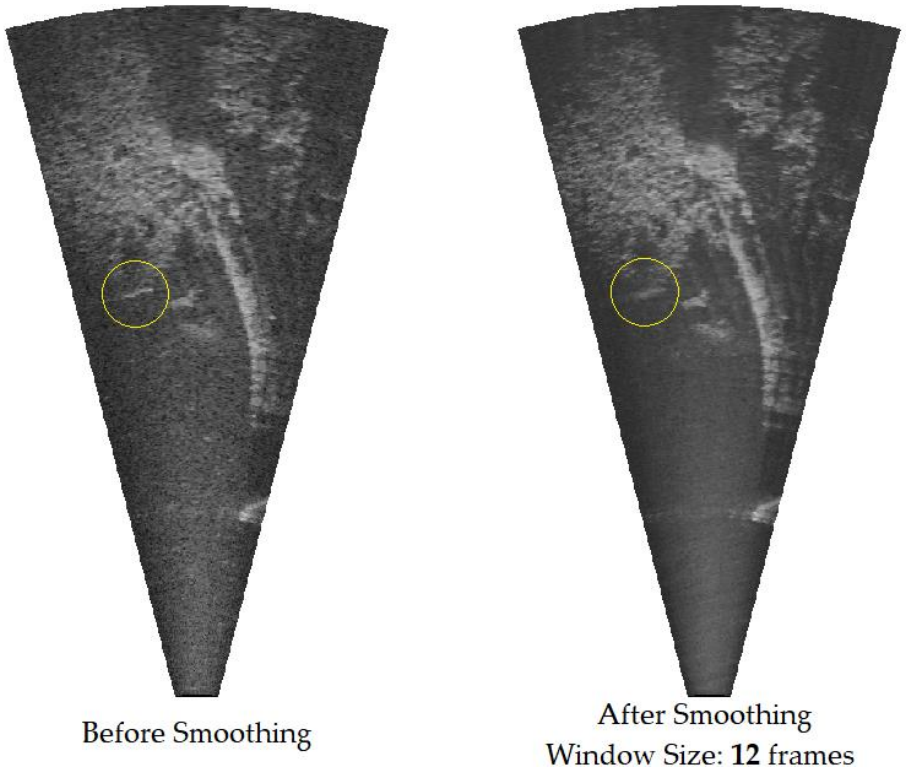

Figure 13. Reconstructed DIDSON image before and after smoothing with a window size equal to 12 frames, with an identified fish target.

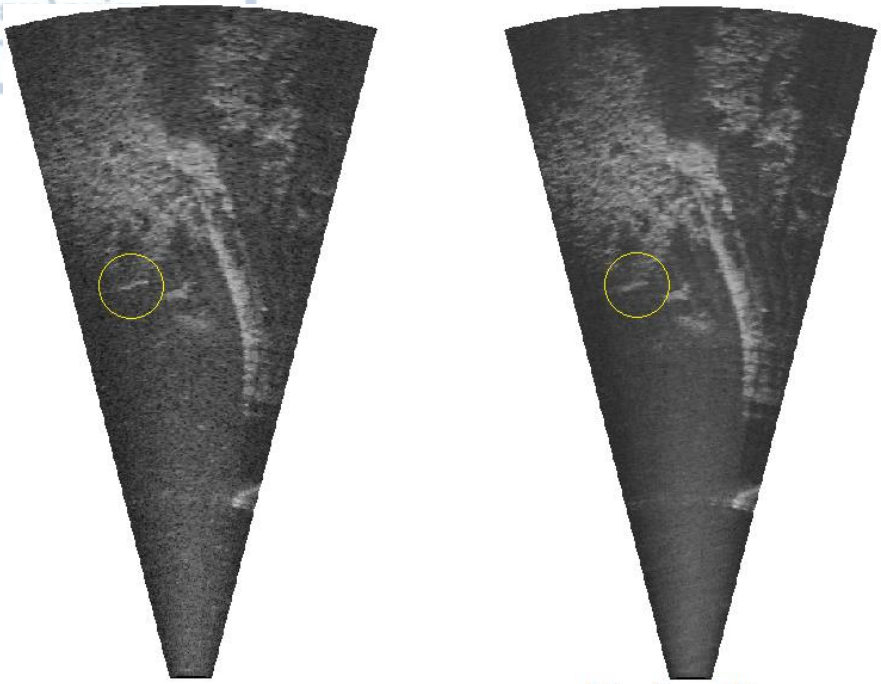

Figure 14. Reconstructed DIDSON image before and after smoothing with a window size equal to 8 frames, with an identified fish target.

3.2.3. Foreground Detection and Extraction

The smoothing process is a vital first step to aid in detecting foreground objects in the reconstructed DIDSON image frame sequence. Numerous algorithmic techniques exist for modelling the background from an image sequence (video), all of which have to deal with the main primary problem, which is the temporal variability in the intensity of the image frames (e.g. Piccardi, 2004; Benezeth et al., 2008). Following the generalization of Benezeth et al. (2008), the general formula that underlies background subtraction methods is:

$$L(i,j,t) = \begin{cases} 1 & if d\big(H(i,j,t), B(i,j)\big) > \tau \\ 0 & otherwise \end{cases}, \qquad (\mathbf{9})$$

where (i, j), represent the pixel index and t represents time, in the image sequence. $I(i, j, t)$ represents the value of the sequence at pixel (i,j) and time t, whereas $B(i, j)$ represents the background value at pixel (i, j). The function $d(I, B)$ is a metric and τ is a threshold value.

The various available methods, then, attempt to model the background *B* and determine a specific metric *d* (Benezeth et al., 2008), whereas the threshold is application-specific.

It is important to note that the separation of an image sequence in foreground and background is artificial and is, in fact, more like an educated hypothesis based on the circumstances, under which the image sequence was acquired. In the case of the DIDSON raw data in the sample used for this thesis, the instrument position was stationary, as it was fixed with respect to the surroundings and the ensonified space at all times during the sequence. For this reason, one potential background model to use would consist of the frame resulting from the average of a brief frame sequence where no target would be detected. Despite the possibility and ease of this method, due to the reason that the methodology mostly aims at automating the procedure, this model would have to rely on manually determining a small frame sequence, where no identifiable fish targets exist. For that reason, an alternative background model was used, which is equally simple to apply. In this choice, the background of each frame was assumed to be the specific frame corresponding to it based on a predetermined time difference t_{lag}:

$$B(i,j,t) = H\big(i,j,t - t_{lag}\big). \qquad (10)$$

This simplifies the foreground calculation as:

$$F(i,j,t) = H(i,j,t) - B(i,j,t) = H(i,j,t) - H\big(i,j,t - t_{lag}\big). \qquad (11)$$

In this work, the value used was t_{lag} = 8. This parameter is, once again, significantly influenced by the velocity distribution of the targets captured in the image sequence and, in turn, significantly influences the clarity and reliability of the resulting foreground. In short, frames located at a temporal distance of t_{lag} apart must be adequately different, in order to reveal a full-body displacement of the target. If targets are moving so slow that they do not cover the complete distance, only part of their body will be identified as a moving image component. To make this point clearer, assume the path followed by a target:

$$P_x(t) = \big(i_{P_0}, j_{P_0}\big) + \big(V_x, V_y\big) \cdot t, \quad (12)$$

where $P_x(t)$ represents the path of a specific pixel located on the edge of a moving target *x* in the sequence, with initial position $P_x(t_0) = (i_{P_0}, j_{P_0})$. For the sake of simplicity, the target is assumed to be rigid, travelling along a linear path, which is oriented along its maximal dimensional axis. All these assumptions are relatively reasonable and are expected to hold for any limited timespan, while typical fish swim primarily along their longest axes, in terms of shape. At a time $t_1 = t_0 + \Delta t$, the position of the given pixel will be:

$$P_x(t_1) = (i_{t_1}, j_{t_1}) = (i_{P_0} + V_x \cdot \Delta t, j_{P_0} + V_y \cdot \Delta t), \quad (13)$$

It is furthermore assumed that the target is contained within the rectangular dimensional range of the depicted image sequence for the entirety of the time span under analysis, i.e. the target does not move out of the visible area. Based on the assumptions that the target is rigid and travelling along their largest axis, the distance between the two pixels is approximately equal to $\sqrt{(V_x \cdot \Delta t)^2 + (V_y \cdot \Delta t)^2}$. If this distance is not at least as large as the dimension of the fish target along its large axis, the foreground resulting from background subtraction using the given time span will lead to the detection of a distorted (smaller, given the current assumptions) target shape. The analysis can be more clearly seen in the simplified diagram of Figure 15 below, where the foreground resulting from the subtraction of the frame at t_2 from the same frame at t_1 will result in target 1 being depicted distorted, with an approximate length equal to $(L_1 - d_1)$.

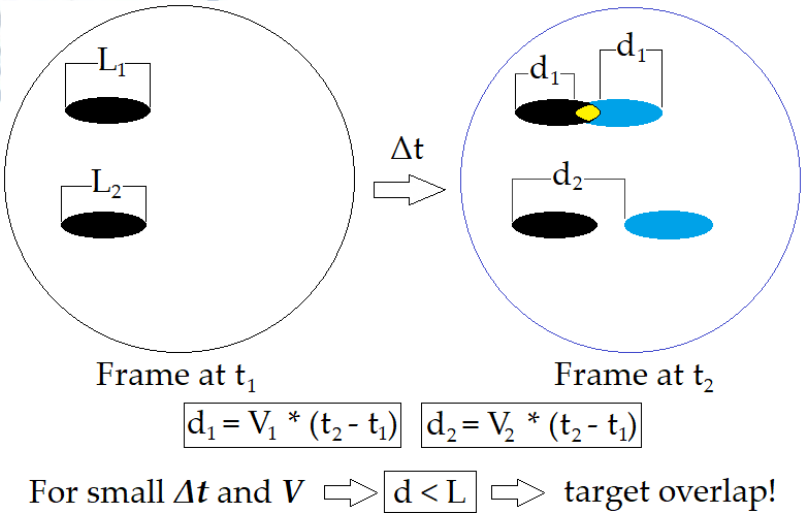

Figure 15. Two frames at times $t_2 > t_1$ depicting moving targets. Depending on target velocity and temporal frame distance, background modelling based on lagged frame subtraction may lead to underestimation of foreground target size (case of target 1 on Figure).

An important note at this point that fish targets are represented on the reconstructed DIDSON image frames by high intensity values, due to stronger reflection and backscatter. Empty interstitial space, on the other hand, is represented by low values, due to lower reflection and backscatter. As a result, if part of a frame F_2 is void, while the corresponding part on a preceding frame F_1 contains a target would result in negative intensities for the subtracted frame $F = F_2 – F_1$ for that location. To avoid "ghost" traces of past targets in future frames, these negative intensity values are converted to zero. Figure 16 depicts a frame and its background model, based on the two aforementioned methods (constant average frame subtraction and lagged frame subtraction). For illustration purposes, a reconstructed DIDSON image frame with a manually identified target was chosen and the target is highlighted on each frame.

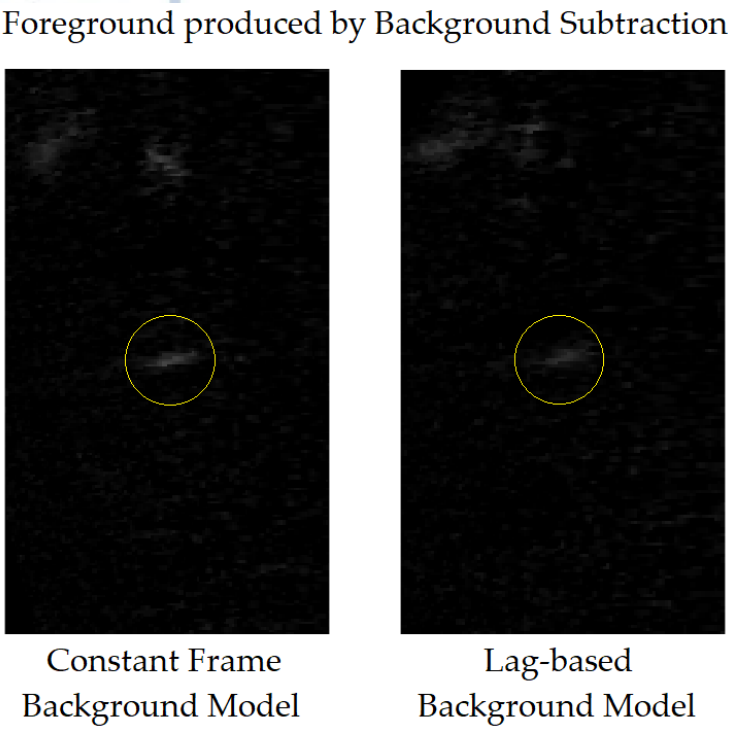

Figure 16. Frame foreground produced through Background Subtraction using two different methods (constant frame model and lag-based model).

The two backgrounds reveal that the constant-frame based model results in a relatively stronger foreground signal (intensity). However, as stated earlier, the background frame was modelled manually, by observing the image sequence and choosing a suitable background frame which is, to the extent possible, empty of targets.

3.2.4. Foreground Masking

The process of foreground extraction is a significant step, which ensures that the effects of motion on the sequence can be studied in isolation from its static parts and is a traditional processing milestone for target detection and tracking methodologies. Given the foreground image, masking techniques can be applied to separate target pixels from the rest of the image frame and isolate them for further analysis (such as counting and tracking). In practice, masking is simply a pixel classification process restricted to 2 classes. Because of this, masking is often carried out by *thresholding*, which is the use of a specific "separator" intensity value to distinguish object pixels from the background. This value can be any one among the values of the intensity range of the frames and, to complicate the matter, as each frame of the sequence has different visual characteristics

and represents a slightly different setting, each frame may require a different threshold for optimal object detection.

A variety of algorithms exist for automatically determining a threshold value that is suitable for object detection on a specific image frame, with many of them using information pertaining to the histogram of the image frame. One of these algorithms that appears to be popular in computer vision applications and which is used in the frame of this work, is *Otsu's method*, which is based on minimizing the intra-class intensity variance (Otsu, 1979). This method performs acceptably well when histograms exhibit *bimodality*, meaning two distinct "peaks" separated by a deep "valley" in between. However, as can be seen in Figure 12 that depicts a histogram representative of the typically available data in the usual scenarios involved in DIDSON-data-based fish target detection, bimodality is a rather unsuitable assumption for the image frame sequence histograms. One important reason for this is that objects are very small, compared to the total background area, in terms of pixels. Apart from small object size, other factors that degrade the quality of thresholding results using Otsu's method are, potentially, the small difference in average values between foreground and background pixels, large variance values of object pixels and/or background pixels, the limited foreground-to-background contrast and the effects of noise (Kittler and Illingworth, 1985; Lee et al., 1990).

After having mitigated the effects of noise through smoothing (pre-processing stage) and increasing the foreground-to-background pixel contrast (foreground extraction stage), as described in the previous sub-chapters, a significant additional improvement is to mitigate the effects of the small object-to-background size ratio prior to applying Otsu's method for thresholding. The technique employed in this work is primarily focused on significantly reducing the number of background pixels on the frame by excluding areas with no apparent motion, as well as pixels that obviously belong to the background. This way, thresholding can be carried out in the regions constrained to an area around the moving targets, which is proportionately smaller, so that the total object-to-background pixel size ratio becomes significantly larger and the histograms of the detected regions-of-motion begin to better approximate bimodality. By applying Otsu's method focusing on the smaller "moving" regions of each image frame of the sequence, improved results can be achieved with respect to target detection and tracking. An example of how constraining the area to threshold around a specific region encompassing a moving target affects the corresponding histogram can be seen in Figure 17. In addition

to limiting the thresholded region, the histogram bin with pixel values equal to 0 has also been excluded from the calculations, in order to intensify the distribution of the rest of the values, as 0 can definitively be considered as background.

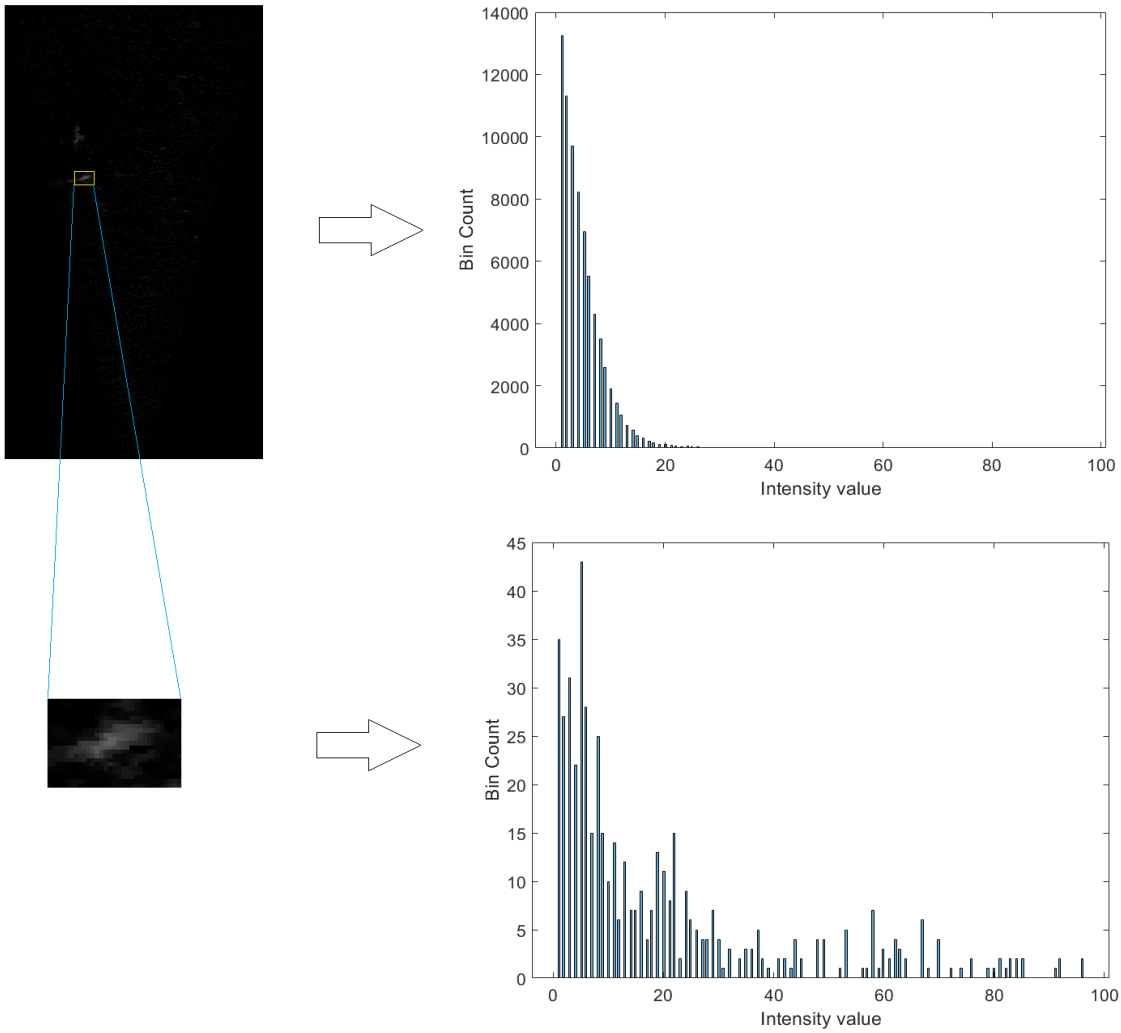

Figure 17. Comparison of histogram of entire image frame to histogram of a small region encompassing an identified target object (pixel values with intensity equal to 0 are excluded from the calculations).

3.2.5. Optical Flow Field Calculation

In order to significantly confine the areas, at which thresholding is to be applied, the optical flow field is used. The calculation follows the MATLAB® scripting environment's implementation of the Farnebäck algorithm. A number of parameters affect the calculation of the optical flow field when using Farnebäck's algorithm, which stem from the intricacies of the algorithm itself. Some of the most important are:

- The number of scales to use for the multi-scale optical flow component estimation (pyramid levels).
- The down-sampling factor between scale levels for the scales used in the iterative calculation (pyramid scale).
- The typical size of each neighborhood that is polynomially-approximated at each step, in pixels.
- The size of the Gaussian filter used to average displacement values estimated from different iterations, in pixels.

Optical flow can be calculated either as a velocity field or as a displacement field. In the frame of this thesis, the two concepts are interchangeable, considering the very small time span between frames, as this is the minimum unit of temporal resolution. For the analysis, the field of velocity magnitudes was used, i.e. the value of $\sqrt{{v_x}^2 + {v_y}^2}$ at each point. For the application of the algorithm in the analysis, three scales were used with a down-sampling factor equal to 0.5, i.e. resolution was doubled at each level. The filter size and neighborhood size were not specifically chosen but were, instead, used as inputs for the genetic algorithm in the process that is described in the next chapter.

The optical flow field calculation step was found to have a profound impact on the quality of the result. Figure 18 below depicts a smooth image frame, along with the smooth foreground (depicted with an equalized histogram), while Figure # depicts the optical flow displacements as calculated on that frame using two different parameter sets.

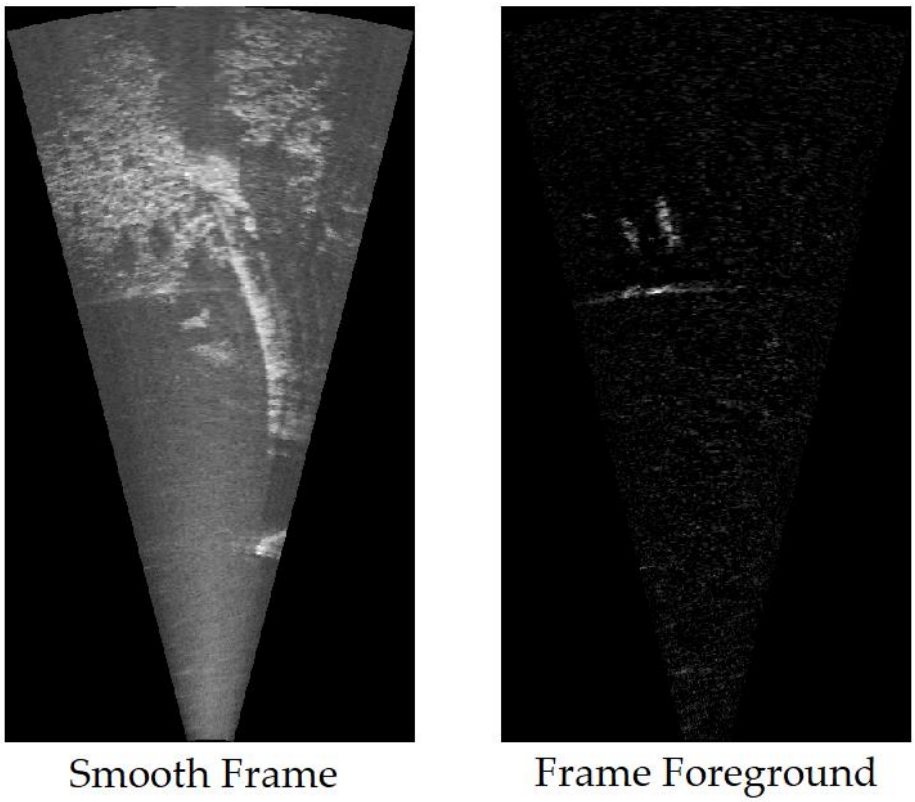

Figure 18. Smooth frame and extracted foreground based on the processes analyzed in the previous subchapters.

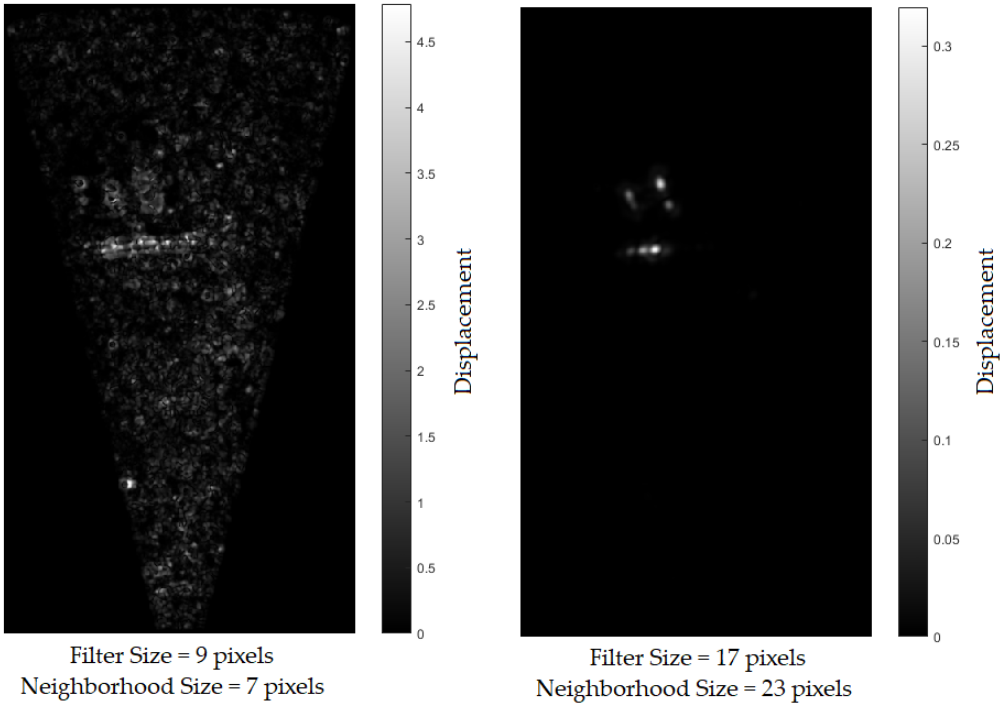

Figure 19. Optical flow displacement field for the image frame of Figure 18, using two different choices for the filter size and the neighborhood size parameters.

3.2.6. Genetic Algorithm – Conditionally Optimal Mask

The filter size and the neighborhood size parameters significantly affect the optical flow field calculation result. The resulting field depends on these parameters in a nonlinear manner, while no straightforward method exists to accurately quantify this dependency. As a consequence, no reliable way exists to choose suitable values for these parameters, beyond experience and testing. Furthermore, their values are constrained to be integers in the context of image processing, as they, in fact, represent pixel units. For this reason, values for these parameters were determined by use of a constrained genetic algorithm, where the solution space is bounded and constrained to the integers. The typically used setup employed the following options:

- 3 pixels < Filter size < 70 pixels.
- 3 pixels < Neighborhood size < 70 pixels.
- 6 individuals (population size).
- 5 generation limit.
- Penalty function 1: Average number of masked pixels per frame.
- Penalty function 2: Constant penalty per very small or very large object.

Two option sets were used, each one using one of the two penalty functions. By using the average number of pixels per frame as the penalty function, the objective of the genetic algorithm is to produce a filter size – neighborhood size parameter pair, which leads to the tightest possible average mask per frame. In that sense, the mask will contain the fewest mask pixels possible, hence avoiding a mask that will be affected by large objects, such as shadows or irregularly dispersed shapes. Furthermore, in order to avoid triggering a convergence to unreasonably low masked pixel counts (for example a totally empty mask), a lower bound is set for both the Filter size and the Neighborhood size, which is chosen to three pixels.

A different condition is used by the second penalty function, namely that detected objects should not bee too small or too large and not too many in total. In specific, objects smaller than 10 pixels in total apply a penalty that is inversely proportional to their size. Similarly, for each object larger than 5000 pixels in total, a very large constant penalty is

applied. This heuristic is expected to lead to less noise, by penalizing too small objects, while practically eliminating very large objects, which are expected only in cases of very unsuitable processing parameters. These two specific penalty functions are expected to lead to results that are *conditionally* optimal.

3.2.7. Output Evaluation

The resulting mask was evaluated by manually counting targets on the 1000-frame reconstructed DIDSON sequence and performing a customized automatic detection of connected components on all frames of the corresponding masked image sequences from the two different solutions. The automated detection was set to include only components with a total number of pixels between 50 and 350, while excluding smaller or larger components. Detected components on each frame were compared with the actual fish targets manually detected on the corresponding frames of the original video. Success rates were calculated for each method, as well as false detections.

4. Results

The dataset used for the analysis was reconstructed to a video sequence of [717 pixel * 400 pixel]-sized image frames and a subset of 1000 frames was kept for the analysis. The genetic algorithm was set up to evolve 5 generations of 6 individual solutions (optical flow field estimation parameters). Using the penalty function of average mask pixels per frame, Figures 20 and 21 below depict the results of a single test run over 5 iterations. Figure 20 depicts the average and the best population penalty at each iteration. Figure 21 depicts the penalties of each generation's solution aggregate. The total intermediate results of the process are shown in Table 3.

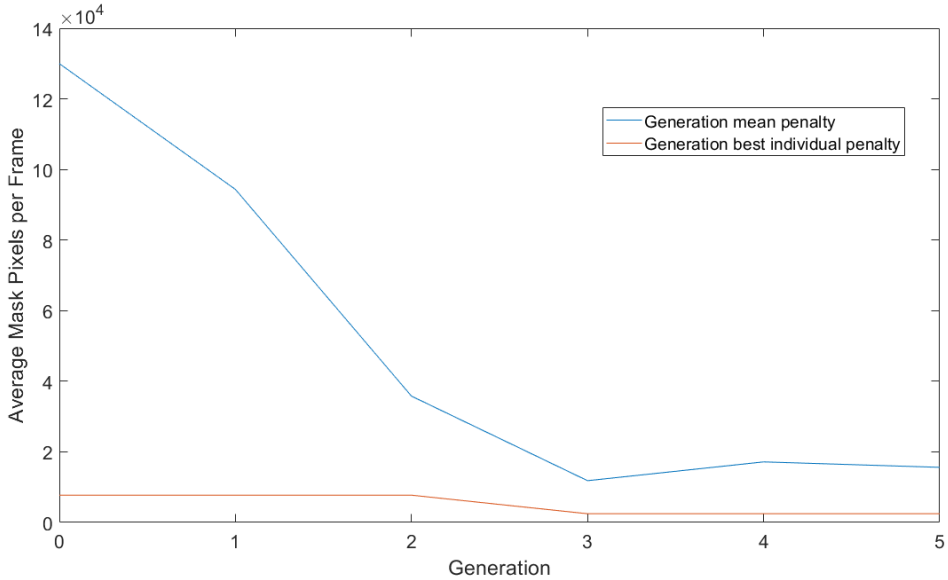

Figure 20. Evolution of the mean and average penalty per generation along a test run of 5 iterations using a penalty function of the average mask pixel count per frame.

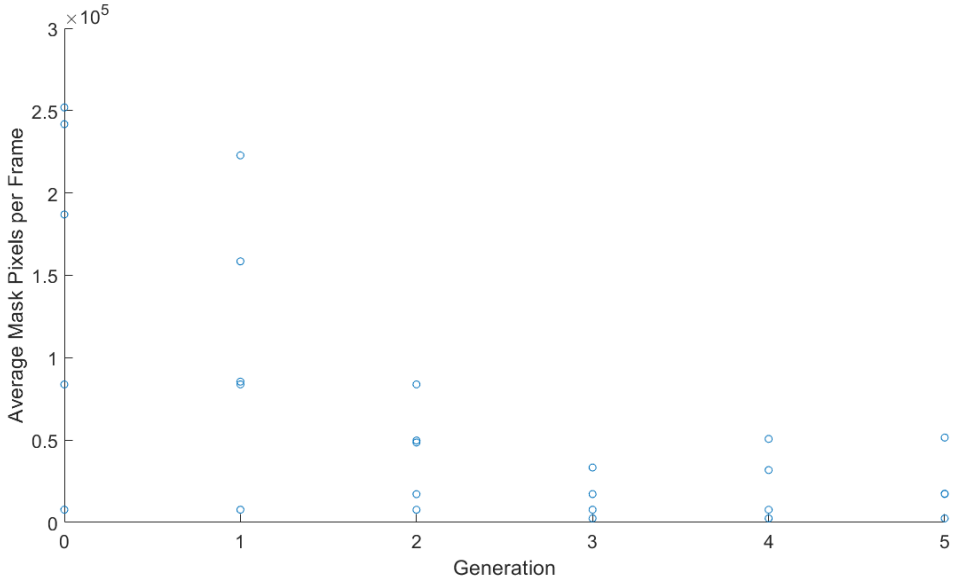

Figure 21. Solution aggregate assessment (penalties) per generation for a test run of 5 iterations using a penalty function of the average mask pixel count per frame.

Table 3. Detailed genetic algorithm results – Solutions and corresponding penalty values for a test run of 5 iterations.

Generation	Filter Size	Neighborhood Size	Average Mask pixels per Frame
1	9	5	83790
	15	10	7729
	23	32	186945
	47	45	241782
	65	49	251938
	15	10	7729
2	9	5	83790
	15	10	7729
	19	31	158501
	13	40	222835
	7	5	85475
	15	10	7729
3	15	10	7729
	9	5	83790

	6	7	48541
	8	7	49768
3	15	9	17157
	15	10	7729
4	15	10	7729
	15	8	33365
	7	9	17234
	5	12	2473
	14	10	7729
	5	12	2473
5	15	10	7729
	14	10	7729
	8	8	31806
	5	12	2473
	11	7	50719
	5	12	2473
6	5	12	2473
	8	9	17460
	3	9	17280
	13	7	51535
	4	12	2473
	5	12	2473

Figures 20 and 21, as well as Table 2, reveal how each successive generation retains the most successful individuals from the previous generations, often in more than 1 copies, while combining individuals and achieving increasingly successful solutions. Based on the genetic algorithm results, the optimal solution was chosen as a Filter Size of 5 pixels and a Neighborhood Size of 12 pixels. Using the same parameters for the case of the second penalty function, which applies a significant penalty for very small and very large objects, leads to the corresponding results shown in Figures 22 and 23.

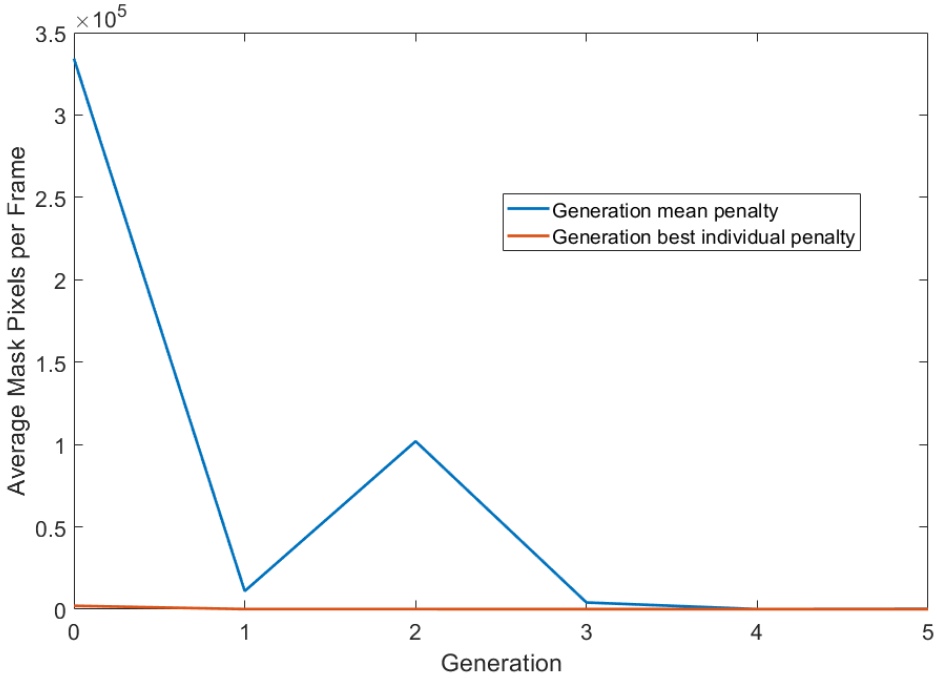

Figure 22. Evolution of the mean and average penalty per generation along a test run of 5 iterations using a penalty function assigning a significant penalty to very large objects and a size-dependent penalty to very small objects.

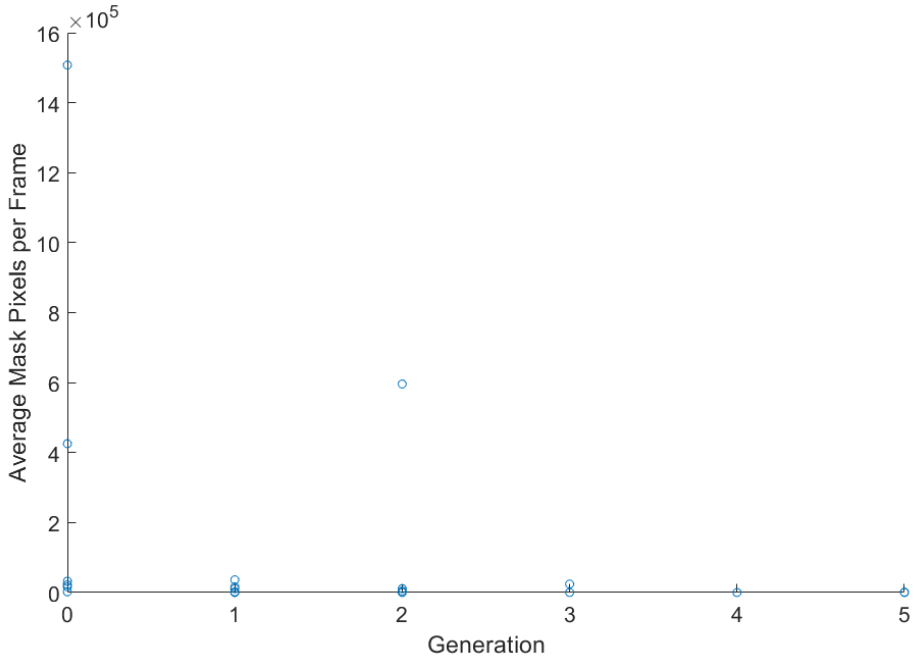

Figure 23. Solution aggregate assessment (penalties) per generation for a test run of 5 iterations using a penalty function assigning a significant penalty to very large objects and size-dependent penalty to very small objects.

The figures reveal that the individual solutions belong to two groups, which are characterized by a very large difference in evaluated penalties, by multiple orders of magnitude, similar to the difference in applied penalties. The ones with the very large values are those that contain one or more very large objects in their masks, which is interpreted as a probable effect of artifacts due to the specific parameters used in the processing. Among the individual solutions with a small penalty, most are kept between generations and, eventually, the mean penalty per generation shows a dramatic reduction, which means that the algorithm converges in the vicinity of solutions that do not produce very large components in the masked frames. The specific penalty values and the filter and neighborhood sizes associated with each individual solution can be seen in Table 4.

Table 4. Detailed genetic algorithm results – Solutions and corresponding penalty values for a test run of 5 iterations using a penalty function assigning a significant penalty to very large objects and size-dependent penalty to very small objects.

Generation	Filter Size	Neighborhood Size	Total Penalty
1	9	5	1507555
	15	10	16252
	19	27	425425
	25	19	32308
	36	12	2107
	29	17	22228
2	36	12	2107
	15	10	16252
	27	12	267
	27	16	11154
	36	18	36307
	33	13	95
3	33	13	95
	27	12	267
	27	16	11154
	31	8	595990

3	25	15	4152
	33	13	95
4	33	13	95
	33	13	95
	33	13	95
	33	13	95
	38	16	24221
	33	14	76
5	33	14	76
	33	13	95
	33	13	95
	33	13	95
	33	13	95
	33	13	95
6	33	14	76
	33	13	95
	33	14	76
	33	13	95
	33	12	1126
	33	14	76

The figure below depicts an image frame that has been masked using the proposed algorithm, based on the conditionally optimal parameter values for the calculation of the optical flow field, with the condition of a minimized mask pixel count per frame.

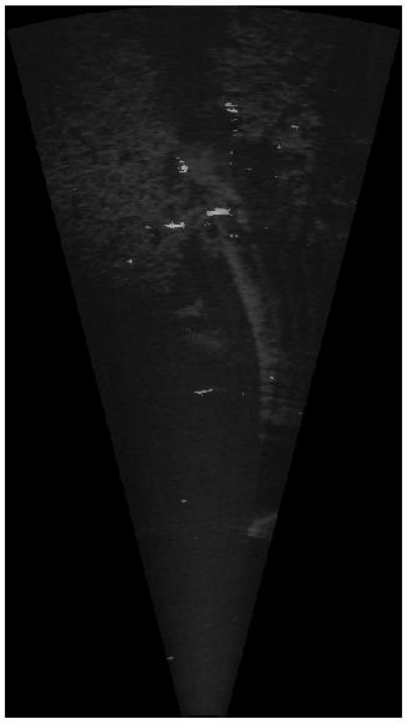

Figure 24. Refined mask calculated using the proposed algorithm, with the conditionally-optimal optical flow parameter values, superimposed on a single smoothed DIDSON frame.

Per-frame target detection and comparison with ground truth was carried out by placing the original image sequence next to the masked image sequence with detected objects and manually evaluating successful and false detections for each frame. A specific frame can be seen for the two cases in Figures 25 and 26. The Figures reveal a cleaner output result using the second penalty evaluation function, which places a penalty to both very small and very large objects. Per-frame target detection success rates were classified for each method in 7 classes, ranging from 0% to 100% and per-frame mis-detections were classified in 12 classes, ranging from 0 to 10+ mis-detected targets. The results can be seen in Figures 27 and 28.

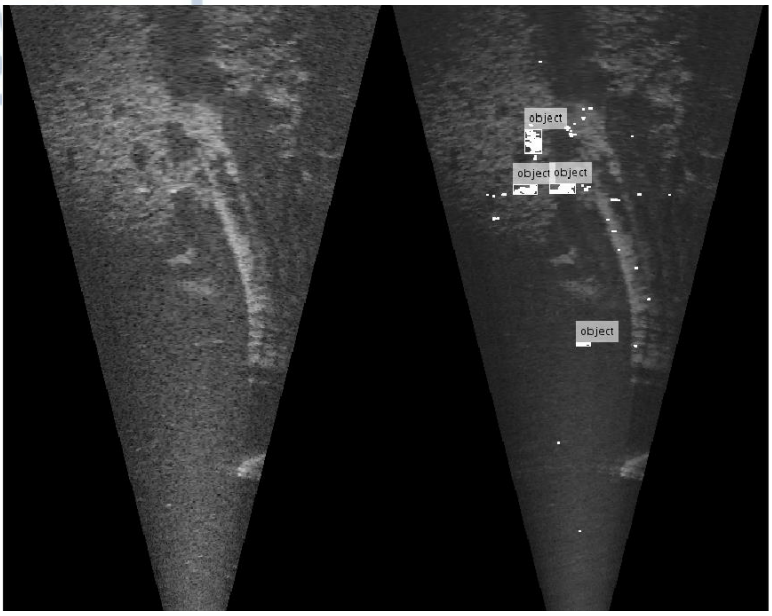

Figure 25. Manual target detection layout frame for result produced using a penalty function of the average mask pixel count per frame.

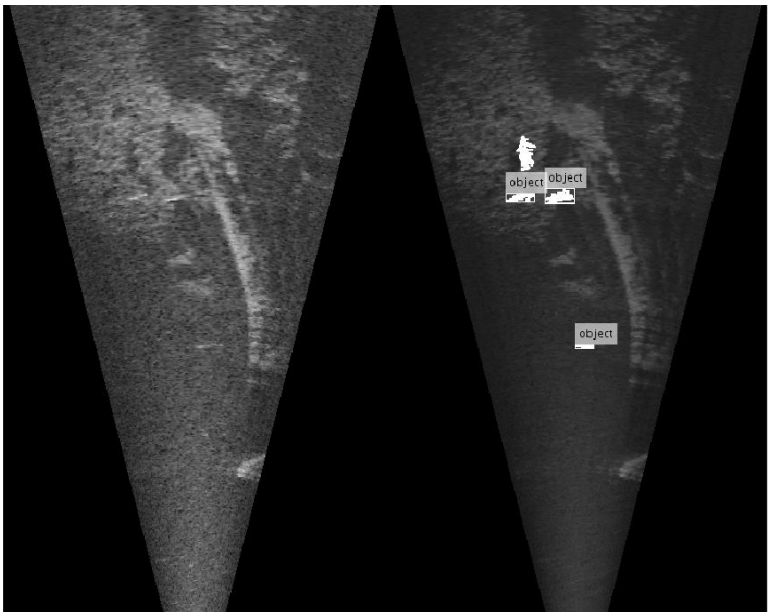

Figure 26. Manual target detection layout frame for result produced using a penalty function assigning a significant penalty to very large objects and size-dependent penalty to very small objects.

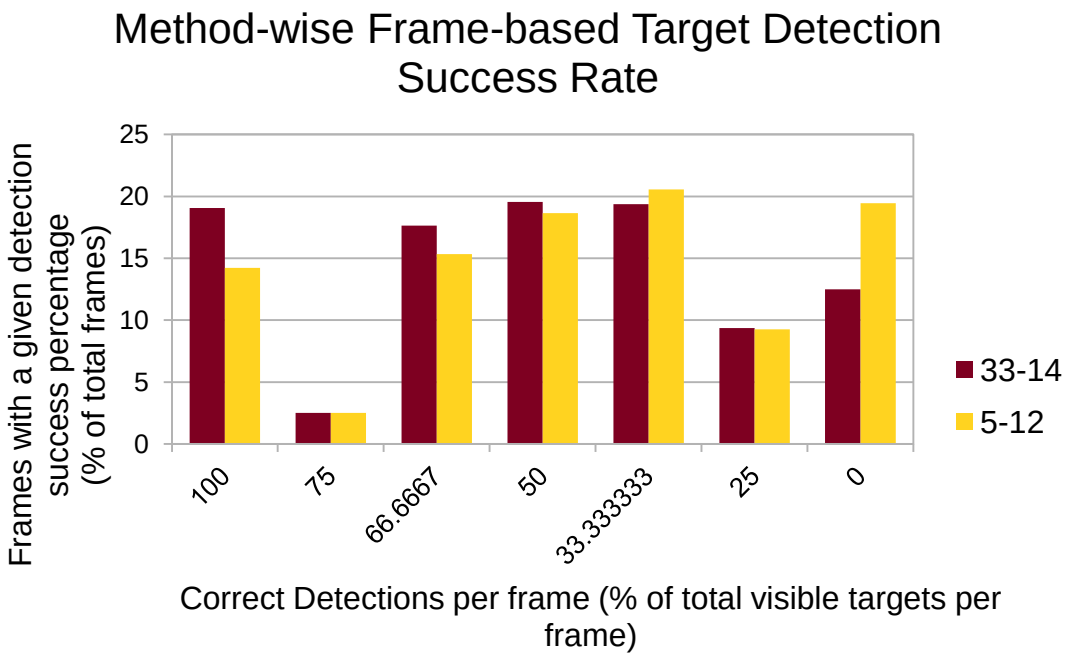

Figure 27. Classified per-frame target detection success rate – Comparison between methods.

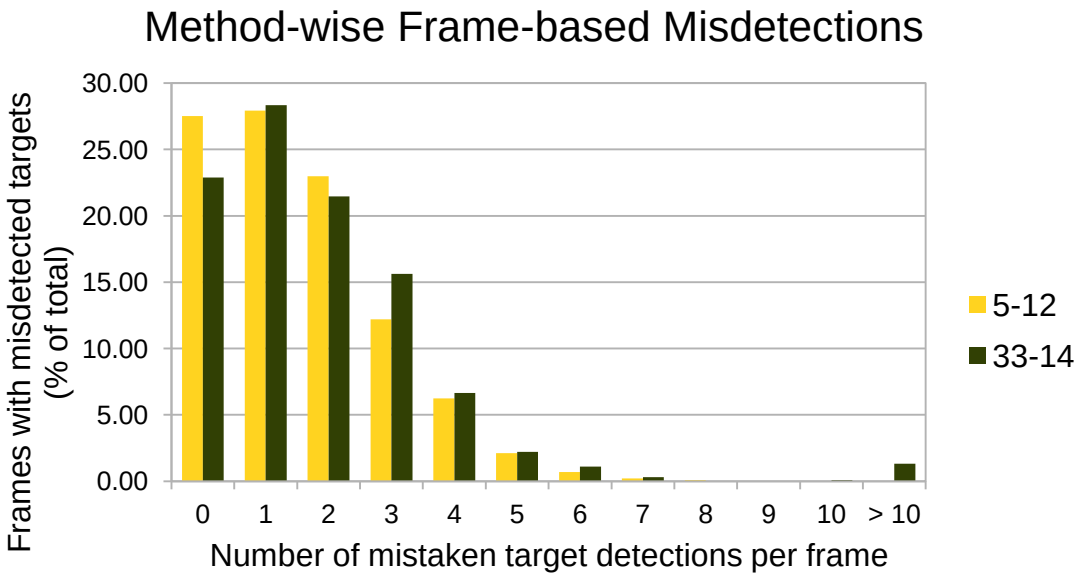

Figure 28. Classified per-frame target mis-detections – Comparison between methods.

Figure 27 indicates that the "33-14" method (filter size of 33 pixels and neighborhood size of 14 pixels) has a generally higher success rate in all success rate

classes, compared to the “5-12” method (filter size of 5 pixels and neighborhood size of 12 pixels). Figure 28, on the other hand, indicates that the “33-14” method results in more per-frame target mis-detections, faring worse than the “5-12” method. While this seems to contradict Figures 25 and 26, which show a cleaner mask for the “33-14” method than the “5-12” method, this result is explained by the detection method, which was set to exclude target detections with a total overall size < 50 pixels. In the final mask of the “5-12” method, such small components are more abundant than in the final mask of the “33-14” method. If these small components had been included by the detection method the result would have been a significantly larger number of target mis-detections for the “5-12” method and a more favourable result profile for the “33-14” method in Figure 28.

A numeric comparison indicating the performance relation between the two methods for all frames is seen in Table 5, which reveals that the “33-14” method has an apparent performance advantage in correct detections.

Table 5. Direct comparison between methods. Indicates the number of frames, for which each method outperformed the other one in terms of correct detections, or the two methods resulted in an overall equal success rate.

Performance Comparison	**Percentage of Frames**
33-14 > 5-12	22.3 %
33-14 > 5-12	11.8 %
33-14 = 5-12	65.9 %

5. Conclusions

This thesis made an attempt to produce locally optimal results through the use of an automated, parametrically-optimized, optical-flow-based, adapted motion detection workflow. The proposed methodology aims to minimize user intervention by employing automated and self-adjusted algorithmic procedures, such as a genetic algorithm. One of the very important steps of the process is the calibration of this genetic algorithm, in order to achieve conditional optimality, by choosing a penalty (or fitness) evaluation function for minimization (or maximization). The chosen penalty functions were, in turn, a) the total masked pixels per frame and b) a weighted sum of the total number of too small and too large distinct connected components per frame for the optical flow field mask. This part of the methodology was employed to overcome the problem of selecting a subset of the parameters, which have a relatively unforeseeable impact on the result.

The key parameters of **filter size** and **neighborhood size** for the step of the calculation of the optical flow field from the DIDSON image sequence were confirmed to significantly affect the finally calculated mask. A [filter size, neighborhood size] choice of [5, 12] appears to conditionally minimize the total masked pixels for each frame, while a choice of [33, 14] conditionally minimizes the total number of distinct connected components. The overall algorithmic procedure has been observed to successfully lead to a locally optimal fish target mask from DIDSON data, producing promising results with minimal user intervention. However, the algorithm strongly depends on a good choice of a *penalty function*, which is not always intuitive and may need to mimic intelligent criteria in a deterministic manner.

Another observation that is worth mentioning is that the two different penalty functions lead to a convergence to almost the same neighborhood size as an input to the optical flow algorithm, namely 12-14 pixels. This is the average size of the region, where motion can be optimally detected, which, in turn, can be interpreted as the largest potential size of the observed moving structures (i.e. fish targets). Depending on the orientation of the targets, their length can be anywhere between 12 to $14\sqrt{2}$ pixels. Since the analyzed reconstructed videos have a total width of 400 pixels, a range of 10 m and a total field of view of 29^{o} correspond to an approximate total width of ~5 m, meaning a total of 5/400 = 1.25 cm pixel width. Therefore, the detected target average length ranges

between 15 cm and ~24.7 cm. These can be taken as indicative limits for the observed fish targets.

While the algorithmic process was used with adequate success in two different datasets, the results need external validation by comparing with alternative methodologies that produce similar outputs. One of the most important drawbacks of the process is the relatively long time taken to calculate the optical flow field calculation algorithm.

Because of the relatively small (and typically bounded) total image sizes in the typical reconstructed DIDSON image sequences, as well as the significant difference in computational overhead between most sub-procedures and the optical flow calculation, the latter dominates the total computational complexity. The process of smoothing is the only sub-process that is as computationally cumbersome as the optical flow field, but is only performed once. As a result, the entire algorithmic process depends primarily on the total individuals and iterated generations of the genetic algorithm, as each generation requires the extraction of the optical flow field (in order to mask it and evaluate the result) for each solution among the individuals. A potential improvement that can be employed to increase the efficiency would be to use parallel computation for the calculation and evaluation of each separate solution individual.

Last but not least, another expected benefit of the present methodology is the possibility of establishing representative optical flow field calculation parameter values for the analysis of DIDSON echosounder-obtained image sequences physically and structurally similar to the dataset sample analyzed in the present thesis. The determination of similarity in this context is, however, an additional problem that can be dealt with in further research efforts.

Bibliography

No. P14V00000374 Soužití člověka a perlorodky říční ve Vltavském luhu, Operational Programme Environment

Alqaddafi, S. (2019). Study the Migration Behavior of Fish by Optical Flow. In *EPH – International Journal of Applied Science* (*ISSN: 2208-2182)*, Vol. 1, No. 1, pp. 709–717.

Balk, H. and T. Lindem (2002). "Fish Detection in Rivers with Split-beam Sonars". In *Proceedings of the 25th Scandinavian Symposium on Physics Acoustics*, U.R. Kristiansen (Ed.), *Ustaoset, Norway, 27-30 January*, Department of Telecommunications, Norwegian University of Science and Technology, pp 6]. [DOI: *Untraced*].

Balk, H., T. Lindem and J. Kubečka (2009). "New Cubic Cross Filter Detector for Multi-beam Data Recorded with DIDSON Acoustic Camera". In *Proceedings of the 3rd International Conference & Exhibition on Underwater Acoustic Measurements: Technologies and Results, Vol. 3*, J. S. Papadakis and L. Bjørnø (Eds.), 21-26 June, Nafplion, Greece [ISBN: *978-9609888332*], pp. 1461-1468.

Baumgartner, L. J., N. Reynoldson, L. Cameron and J. Stanger (2006). Assessment of a Dual-Frequency Identification Sonar (DIDSON) for Application in Fish Migration Studies. New South Wales (NSW) Department of Primary Industries, Fisheries Final Report Series No. 84, 33 pp. [URI: *https://www.dpi.nsw.gov.au/__data/assets/pdf_file/0009/136665/Output-686_DIDSON-Report_FORMATTED.pdf*], [ISSN: *1449-9967*]

Beauchemin, S. S. and J. L. Barron (1995). The Computation of Optical Flow. In *ACM Computing Surveys*, Vol. 27, No. 3, pp. 433–466. [DOI: *10.1145/212094.212141*]

Becker, A., A. K. Whitfield, P. D. Cowley, J. Järnegren and T. F. Næsje. (2011). An Assessment of the Size Structure, Distribution and Behaviour of Fish Populations within a Temporarily Closed Estuary using Dual Frequency Identification Sonar (DIDSON). In *Journal of Fish Biology*, Vol. 79, No. 3, pp. 761-775. [DOI: *10.1111/j.1095-8649.2011.03057.x*]

Belcher, E. O. (2006) Vision in Turbid Water. Sound Metrics Corp., Lake Forest Park, Washington, U.S.A., 5 pp. [URI: http://www.soundmetrics.com/getmedia/5535cf70-7952-4131-91cc-28e65bbdeef7/UI2007Paper_BELCHER.pdf.aspx]

Belcher, E. O., W. L. J. Fox and W. H. Hanot (2002). "Dual-Frequency Acoustic Camera: A Candidate for an Obstacle Avoidance, Gap-Filler, and Identification Sensor for Untethered Underwater Vehicles". In *Proceedings of the MTS/IEEE Conference: Oceans 2002*, 29-31 October, Biloxi, Michigan, U.S.A., pp. 2124-2128. [DOI: *10.1109/OCEANS.2002.1191959*]

Belcher, E., W. Hanot and J. Burch (2002). "Dual-frequency Identification Sonar (DIDSON)". In *Proceedings of the 2002 International Symposium on Underwater Technology*: UT 02, 16-19 April, Tokyo, Japan, pp. 187-192. [DOI: *10.1109/UT.2002.1002424*]

Belcher, E., B. Matsuyama and G. Trimble (2001). "Object Identification with Acoustic Lenses". In *Proceedings of the MTS/IEEE Conference: Oceans 2001 – An Ocean Odyssey*, 5-8 November, Honolulu, Hawaii, U.S.A. (IEEE Cat. No. 01CH37295), Vol. 1, pp. 6-11. [DOI: *10.1109/OCEANS.2001.968656*]

Benezeth, Y., P. M. Jodoin, B. Emile, H. Laurent and C. Rosenberger (2008). "Review and Evaluation of Commonly-Implemented Background Subtraction Algorithms". In *Proceedings of the 19th International Conference on Pattern Recognition*, December 8–11, Tampa, Florida, U.S.A., 4 p. [DOI: *10.1109/ICPR.2008.4760998*]

Bertero, M., T. A. Poggio and V. Torre (1988). "Ill-posed Problems in Early Vision". In *Proceedings of the IEEE*, Vol. 76, No. 8, pp 869–889. [DOI: *10.1109/5.5962*]

Bevelhimer, M. S., C. Scherelis, J. Colby, C. Tomichek and M. Adonizio (2015). "Fish Behavioral Response during Hydrokinetic Turbine Encounters Based on Multi-beam Hydroacoustics Results". In *Proceedings of the 3rd Marine Energy Technology Symposium (METS)*, 27-29 April, Washington, DC, U.S.A. [URI: *https://tethys.pnnl.gov/sites/default/files/publications/Bevelhimer-et-al-2015-METS.pdf*]

Bruhn, A., J. Weickert and C. Schnörr (2005). Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic Flow Methods. In *International Journal of Computer Vision*, Vol. 61, No. 3, pp. 211–231. [DOI: *10.1023/B:VISI.0000045324.43199.43*]

Bulas-Cruz, J., A. T. Ali and E. L. Dagless (1993). "A Temporal Smoothing Technique for Real-Time Motion Detection". In *Proceedings of the International Conference on Computer Analysis of Images and Patterns*, September 13–15, Budapest, Hungary, pp. 379–386. *Lecture Notes in Computer Science*, Vol. 719, Springer, Berlin, Heidelberg. [DOI: *10.1007/3-540-57233-3_50*]

Burwen, D. L., S. J. Fleischman and J. D. Miller (2010). Accuracy and Precision of Salmon Length Estimates Taken from DIDSON Sonar Images. In *Transactions of the American Fisheries Society*, Vol. 139, No. 5, pp. 1306-1314. [DOI: *10.1577/T09-173.1*]

De Jong, K. A. (1993). "Genetic Algorithms are NOT Function Optimizers". In *Foundations of Genetic Algorithms*, Vol. 2, pp. 5-17. [DOI: *10.1016/B978-0-08-094832-4.50006-4*]

Elliott, D. B. (2005). DIDSON Diver Evaluation. NOAA Diver – National Ocean Service, Silver Spring, Maryland, U.S.A., 5 pp. [URI: *http://www.soundmetrics.com/getmedia/5ff2a93d-5fe9-4af6-8e2e-75e9ae7fefb3/DIDSON_DH_Evaluation.pdf.aspx*]

Farnebäck, G. (2003). "Two-Frame Motion Estimation Based on Polynomial Expansion". In "Image Analysis", J. Bigun and T. Gustavsson (Eds.), *Scandinavian Conference in Image Analysis*, pp. 363–370. *Lecture Notes in Computer Science*, Vol. 2749, Springer, Berlin, Heidelberg. [DOI: *10.1007/3-540-45103-X_50*]

Faulkner, A. V. and S. L. Maxwell (2009). An Aiming Protocol for Fish-counting Sonars using River Bottom Profiles from a Dual-Frequency Identification Sonar (DIDSON). Fishery Manuscript No. 09-03, Alaska Department of Fish and Game, Division of Sport Fish and Commercial Fisheries, Research and Technical Services. [URI: *http://adfg.alaska.gov/FedAidPDFs/fms09-03.pdf*]

Fernandes, P. G., F. Gerlotto, D. V. Holliday, O. Nakken and E. J. Simmonds (2002). Acoustic Applications In Fisheries Science: The ICES Contribution. In *100 Years of Science under ICES*, Helsinki, 1-4 August, 2000, *International Council for the*

Exploration of the Sea (*ICES*) *Marine Science Symposia*, Vol. 215, pp. 483-492. [DOI: *Untraced*]

Fleet, D. J. and Y. Weiss (2005). Optical Flow Estimation. In *Mathematical Models in Computer Vision: The Handbook*, N. Paragios, Y. Chen and O. Faugeras (Eds.), Springer, Chapter 15, pp. 239–258.

Foote, K. G. (1983). Linearity of Fisheries Acoustics, with Addition Theorems. In *The Journal of the Acoustical Society of America*, Vol. 73, No. 6, pp. 1932-1940. [DOI: *10.1121/1.389583*]

Galbreath, P. F., and P. E. Barber (2005). Validation of a Long-range Dual Frequency Identification Sonar (DIDSON-LR) for Fish Passage Enumeration in the Methow River. *Columbia River Inter-Tribal Fish Commission* (*CRITFC*) *Technical Report #05-04*, November 10, Portland, Oregon, U.S.A. [URI: *https://www.critfc.org/wp-content/uploads/2012/11/05_04report.pdf*]

Godlewska, M., N. Górska, Z. Klusek, J. Szczucka, A. Œwierzowski and J. Têgowski (2002). "Hydroacoustics as a Tool for Monitoring the Quality of Aquatic Ecosystems". In *Ecology and Ecotechnologies, Proceedings of the Review Conference on the Scientific Cooperation between Austria and Poland*, February 24-28, Scientific Centre of the Polish Academy of Sciences, Vienna, Austria, Vol. 2, pp. 287-295.

Hadamard, J. (1902). Sur les Problèmes aux Dérivées Partielles et Leur Signification Physique. In *Princeton University Bulletin*, Vol. 13, pp. 49–52.

Han, J., N. Honda, A. Asada and K. Shibata (2009). Automated Acoustic Method for Counting and Sizing Farmed Fish During Transfer Using DIDSON. In *Fisheries Science*, Vol. 75, No. 6, pp. 1359-1367. [DOI: *10.1007/s12562-009-0162-5*]

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*: *An Introductory Analysis with Applications to Biology*, *Control*, *and Artificial Intelligence*. University of Michigan Press. [ISBN: *978-047208460-9*]

Holmes, J. A., G. M. Cronkite, H. J. Enzenhofer and T. J. Mulligan (2006). Accuracy and Precision of Fish-Count Data from a "Dual-Frequency Identification Sonar" (DIDSON) Imaging System. In *International Council for the Exploration of the Sea* (*ICES*) *Journal of Marine Science*, Vol. 63, No. 3, pp. 543-555. [DOI: *10.1016/j.icesjms.2005.08.015*]

Horn, B. K. P. and B. G. Schunck (1981). Determining Optical Flow. In *Artificial Intelligence*, Vol. 17, No. 1–3, pp. 185–203. [DOI: *10.1016/0004-3702(81)90024-2*]

Hughes, J. B. (2012). *Combining Count Data from Split-beam and Multiple DIDSON Sonar Techniques to Estimate Spawning Run Abundance of Anadromous Fishes in the Roanoke River, NC*. MSc Thesis, North Carolina State University, U.S.A, 9 May. [URI: *http://www.lib.ncsu.edu/resolver/1840.16/7894*]

Husseini, S. (2017). A Survey of Optical Flow Techniques for Object Tracking, *B.Sc. Thesis, Tampere University of Technology*, Finland.

Johannsson, H., M. Kaess, B. Englot, F. Hover and J. Leonard (2010). "Imaging Sonar-Aided Navigation for Autonomous Underwater Harbor". In *Proceedings of the*

IEEE/RJS International Conference on Intelligent Robots and Systems, October 18-22, Taipei, Taiwan, pp. 4396–4403. [DOI: *10.1109/IROS.2010.5650831*]

Johnson, E. L., T. S. Clabough, M. L. Keefer, C. C. Caudill, P. N. Johnson, W. T. Nagy and M. A. Jepson (2012). Evaluation of Dual Frequency Identification Sonar (DIDSON) for Monitoring Pacific Lamprey Passage Behavior at Fishways of Bonneville Dam, 2011. Technical Report No. 2012-5, Idaho Cooperative Fish and Wildlife Research Unit, U.S. Army Corps of Engineers, Portland, U.S.A. [URI: *https://usace.contentdm.oclc.org/digital/collection/p16021coll3/id/80/*]

Kang., M. H. (2011). Semiautomated Analysis of Data from an Imaging Sonar for Fish Counting, Sizing, and Tracking in a Post-Processing Application. *Fisheries and Aquatic Sciences*, Vol. 14, No 3, pp. 218-225. [DOI: *10.5657/FAS.2011.0218*]

Kittler, J. and J. Illingworth (1985). On Threshold Selection Using Clustering Criteria. In *IEEE Transactions On Systems, Man, and Cybernetics*, Vol. SMC-15, No. 5 (September-October), pp. 652-655. [DOI: *10.1109/TSMC.1985.6313443*]

Lankhorst, M. M. (1996). *Genetic Algorithms in Data Analysis*, PhD Thesis, Rijksuniversiteit Groningen (University of Groningen).

Langkau, M. C., Balk, H., Schmidt, M. B. and J. Borcherding, (2012). Can acoustic shadows identify fish species? A novel application of imaging sonar data. *Fisheries Management and Ecology*, Vol.19, No.4, pp. 313-322.

Lee, S. U., S. Y. Chung and R. H. Park (1990). A Comparative Performance Study of Several Global Thresholding Techniques for Segmentation. In *Computer Vision,*

Graphics, and Image Processing, Vol. 52, No. 2, pp. 171-190. [DOI: *10.1016/0734-189X(90)90053-X*]

Li, Y.-X. And M. Gen (1996). "Nonlinear Mixed Integer Programming Problems Using Genetic Algorithm and Penalty Function". In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, October 14–17, Beijing, China, pp. 2677–2682. [DOI: 10.1109/ICSMC.1996.561362]*

Lilja, J. and P. Orell (2011). Use of DIDSON to estimate spawning run of Atlantic salmon in the River Karasjohka, the tributary of the River Tana, Game and Fisheries Research, Norwegian Institute for Nature Research, 25 pp [URI: *https://www.tenojoki.fi/tenopaasivut/arkisto2/final_compiled_report_didson_karasjohka_2010.pdf*]

Lozano, J. A., P. Larrañaga, M. Graña, and F. X. Albizuri (1999). "Genetic Algorithms: Bridging the Convergence Gap". In *Theoretical Computer Science*, Vol. 229, No. 1-2, pp. 11-22. [DOI: *10.1016/S0304-3975(99)00090-0*]

Lucas, B. D. (1984). *Generalized Image Matching by the Method of Differences*. PhD Thesis, School of Computer Science, Carnegie–Mellon University, Pittsburgh, PA.

Lucas, B. and T. Kanade (1981). "An Iterative Image Registration Technique with an Application to Stereo Vision". In Volume 2 of the *Proceedings of the Seventh International Joint Conference on Artificial Intelligence,* August 24-28, Vancouver, Canada, pp. 674–679.

MacLennan, D. N., P. G. Fernandes and J. Dalen (2002). A Consistent Approach to Definitions and Symbols in Fisheries Acoustics. In *ICES (International Council for the Exploration of the Sea) Journal of Marine Science*, Vol. 59, No. 2, pp. 365-369. [DOI: *10.1006/jmsc.2001.1158*]

Martignac, F., A. Daroux, J. L. Bagliniere, D. Ombredane, and J. Guillard (2015). The Use of Acoustic Cameras in Shallow Waters: New Hydroacoustic Tools for Monitoring Migratory Fish Population. A Review of DIDSON Technology. In *Fish and Fisheries*, Vol. 16, No. 3, pp. 486-510. [DOI: *10.1111/faf.12071*]

Maxwell, S. L. and N. E. Gove (2004). The Feasibility of Estimating Migrating Salmon Passage Rates in Turbid Rivers Using a Dual-Frequency Identification Sonar (DIDSON). Regional Information Report No. 2A04-05, Alaska Department of Fish and Game, Division of Commercial Fisheries. [URI: *http://www.adfg.alaska.gov/FedAidPDFs/RIR.2A.2004.05.pdf*]

Maxwell, S. L. and A. V. Smith (2007). Generating River Bottom Profiles with a Dual-Frequency Identification Sonar (DIDSON). In *North American Journal of Fisheries Management*, Vol. 27, No. 4, pp. 1294-1309. [DOI: *10.1577/M07-019.1*]

Mueller, R. P., R. S. Brown, H. Hop and L. Moulton (2006). Video and Acoustic Camera Techniques for Studying Fish Under Ice: A Review and Comparison. In *Reviews in Fish Biology and Fisheries*, Vol. 16, No. 2, pp. 213-226. [DOI: *10.1007/s11160-006-9011-0*]

Mueller, A. M., T. Mulligan and P. K. Withler (2008). Classifying Sonar Images: Can A Computer-Driven Process Identify Eels?. In *North American Journal of Fisheries Management*, Vol. 28, No. 6, pp. 1876-1886. [DOI: *10.1577/M08-033.1*]

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. In *IEEE Transactions on Systems, Man, and Cybernetics,* Vol. 9, No. 1, pp. 62-66. [DOI: *10.1109/TSMC.1979.4310076*]

Petreman, I. C., N. E. Jones and S. W. Milne (2014). Observer Bias and Subsampling Efficiencies for Estimating the Number of Migrating Fish in Rivers Using Dual-frequency IDentification SONar (DIDSON). In *Fisheries Research*, Vol. 155, pp. 160-167. [DOI: *10.1016/j.fishres.2014.03.001*]

Piccardi, M. (2004). "Background Subtraction Techniques: A Review". In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, October 10–13, Hague, Netherlands, pp. 3099–3104. [DOI: *10.1109/ICSMC.2004.1400815*]

Pipal, K., M. Jessop, D. Boughton and P. Adams (2010). Using Dual-Frequency Identification Sonar (DIDSON) to Estimate Adult Steelhead Escapement in the San Lorenzo River, California. *California Fish and Game*, Vol. 96, No. 1, pp. 90-95. [DOI: *Untraced*]

Rakowitz, G., Tušer, M., Říha, M., Jůza, T., Balk, H., and Kubečka, J. (2012). Use of high-frequency imaging sonar (DIDSON) to observe fish behaviour towards a surface trawl. *Fisheries Research,*Vol.123, pp. 37-48.

Reid, D. G. (2000). Report on Echo Trace Classification. International Council for the Exploration of the Sea (ICES) Cooperative Research Report No. 238, Copenhagen, Denmark, 107 pp. [DOI: *10.17895/ices.pub.5371*]

Rudolph, G. (1994). "Convergence Analysis of Canonical Genetic Algorithms". In *IEEE Transactions on Neural Networks*, Vol. 5, No. 1, pp. 96-101. [DOI: *10.1109/72.265964*]

Simmonds, E. J., N. J. Williamson, F. Gerlotto and A. Aglen (1992). Acoustic Survey Design and Analysis Procedure: A Comprehensive Review of Current Practice. International Council for the Exploration of the Sea (ICES) Cooperative Research Report No. 187, Copenhagen, Denmark, 112 pp. [DOI/ISBN: *Untraced*]

Simonoff, J. S. (1998). *Smoothing Methods in Statistics*, Springer-Verlag, New York Berlin Heidelberg, 1996, Corrected Second Edition May 1998. [ISBN: *978-0387947167*]

Ullman, S. (1979). *The Interpretation of Visual Motion*. MIT Press, Cambridge, London, 1979. [ISBN: *978-0262710114*].

Vaganay, J., M. L. Elkins, S. Willcox, F. S. Hover, R. S. Damus, S. Desset, J. P. Morash and V. C. Polidoro (2005). "Ship Hull Inspection by Hull-Relative Navigation and Control". In *Proceedings of MTS/IEEE Conference: Oceans 2005 – One Ocean*, September 18/19-23, Washington DC, U.S.A., pp. 761-766. [DOI: *10.1109/OCEANS.2005.1639844*]

Walsh, S. J., A. Engas, R. Ferro, R. Fonteyne and B. V. Marlen (2002). To Catch or Conserve More Fish: the Evolution of Fishing Technology in Fisheries Science. In *100 Years of Science under ICES*, Helsinki, 1-4 August, 2000, *International Council for the Exploration of the Sea* (*ICES*) *Marine Science Symposia*, Vol. 215, pp. 493-503. [DOI: *Untraced*]

Winfield, I. J. (2004). "Fish Population Ecology". Chapter 15 of *The Lakes Handbook: Limnology and Limnetic Ecology*, Vol. 1, O'Sullivan, P. E., Reynolds, C. S. [Eds.], Oxford, Blackwell Science Limited. [ISBN: *978-0-632-04797-0*]

Wu, N. (2012). Fish Detection in Underwater Video of Benthic Habitats in Virgin Islands. *Open Access Theses*, 348. (*https://scholarlyrepository.miami.edu/oa_theses/348*)