



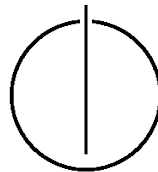
FAKULTÄT FÜR INFORMATIK

DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Mater's Thesis in Biomedical Computing

**The Big Work - Deformable object detection
in underwater imaging**

Andrés Sánchez





FAKULTÄT FÜR INFORMATIK

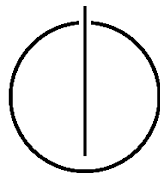
DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Mater's Thesis in Biomedical Computing

The Big Work - Deformable object detection in
underwater imaging

Deformierbare Objekterkennung in Unterwasser-Bilder

Author:	Andrés Sánchez
Examiner:	Prof. Dr. Nassir Navab
Supervisor:	Prof. Dr. Slobodan Ilic
Advisor:	M.Sc. David J. Tan
Date:	November 27, 2013



I hereby declare that this thesis is entirely the result of my own work except where otherwise indicated. I have only used the resources given in the list of references.

München, den 1. Juli 2014

Andrés Sánchez

Acknowledgments

If someone contributed to the thesis... might be good to thank them here.

Abstract

An abstracts abstracts the thesis!

The monitoring of fish for stock assessment in aquaculture, commercial fisheries and in the assessment of the effectiveness of biodiversity management strategies such as marine protected Areas and closed area management has been thriving since the 1980s. as does area continuously grows, it becomes important to develop a remote monitoring system to estimate the biomass of the large number of fishes bred in cages, since around 80% of all sales of farmed fish are arranged pre-harvest, that mean, the profit on the sale directly depends on correct estimations of weight, size distribution and total biomass. Therefore automated and relatively affordable tools for biomass estimation have to be developed.

Here, we will rely on complex stereo camera system, composed of time of flight range camera and CCD grayscale camera, that film fishes in the cage for certain period of time. in order to estimate the biomass, the volume of the fish has to be estimated. this can be achieved by first detecting and segmenting the fish in every grayscale image of the incoming video stream and then translate this found fish contour to the range image obtaining a estimation of the volume. to find the algorithm that is in line with our problem, we need to understand the challenge in detecting fishes. they include the motion of the fish which makes the object of interest deformable, the location of the fish respect to the camera and occlusion caused by having multiple fishes in every available frame.

In this project, we concentrate on the first step that is detection of the fish that undergo deformation in grayscale images. Inspired by recent works in we develop a similar approach for fish detection. we use a

We evaluate the proposed method by computing difference between the label dataset with the predicted result, in addition, we cluster the results from different camera locations and found that when the sagittal plane is parallel to the image plane, the tracking algorithm provide the best result. Finally, we show that .

Therefore, this thesis accomplished the following:

define recent work

define algorithm

show improvement

define accomplish, if there are somethings

Contents

Acknowledgements	vii
Abstract	ix
Outline of the Thesis	xiii
I. Overview	1
1. Introduction	3
1.1. Motivation	3
1.2. Problem Statement	3
2. Related Work	7
2.1. Related Work	7
2.2. Problem Statement	8
II. Methods and Implementation	9
3. Methods	11
3.1. Theoretical Background	11
3.2. Notation and Symbols	11
3.3. Part Based Model	11
3.4. Linemod	11
4. Implementation	13
4.1. Fish Model	13
4.2. Linemod	13
III. Results and Conclusion	15
5. Results and Discussion	17
5.1. Results and Discussion	17
5.2. Conclusion	17
6. Conclusion	19
6.1. Discussion	19
6.2. Conclusion	19

Appendix	23
A. Detailed Descriptions	23
Bibliography	25

Outline of the Thesis

Part I: Overview

CHAPTER 1: INTRODUCTION

This chapter presents an overview of the thesis and its purpose. Furthermore, it will discuss the sense of life in a very general approach.

CHAPTER 2: RELATED WORKS

No thesis without theory.

Part II: Methods

CHAPTER 3: THEORETICAL BACKGROUND

This chapter presents the requirements for the process.

CHAPTER 4: IMPLEMENTATION

This chapter presents the requirements for the process.

Part III: Results and Conclusion

CHAPTER 5: RESULTS AND DISCUSSION

This chapter presents the requirements for the process.

CHAPTER 6: CONCLUSION

This chapter presents the requirements for the process.

Part I.

Overview

1. Introduction

1.1. Motivation

As we progress from livelihood fisheries to aquaculture industries, the global production and demand of fishes has drastically increased over several decades. According to [Asche and Bjørndal \[2011\]](#), the production increased from 16 M in the 1970s to 142 M in 2008. In these statistical figures, the amount of wild fishes has reached a threshold since 1980s while the farmed fishes picked up the difference in amount. For instance, [LARSEN and ASCHE](#) mention in [\[2011\]](#) that Norway alone increased their production of Salmons from a few thousand in the 1980s to approximately 1.4 M in 2009 which constitutes around 51% of the global supply. This makes then the largest supplier of Salmons in the world [[Asche and Bjørndal, 2011](#); [LARSEN and ASCHE, 2011](#); [Liu et al., 2011](#)].

Other than favourable geographical and environmental features that made Norway viable for this industry technological advancement also played an important role in the economical cycle between demand and supply. As production increase, they reduced cost and as a consequence, increased the demand [Asche and Bjørndal \[2011\]](#). Therefore, this cycle supported the growth of the industry over the years. Since around, 80 % of all sales of farmed fish are arranged pre-harvest, the profit on the sale directly depends on the correct estimations of weight, size distribution and total biomass. Therefore, our project deals with remote monitoring of fishes size and weight distribution in aquaculture environments. Considering a large amount of fish, it becomes essential to develop an automated biomass estimator to constantly monitor the changes or growth of fishes. This system involves cameras that would detect the fish in a video sequence and compute the biomass distribution over a specified period of time.

Add description of hardware setup

check this paragraph

1.2. Problem Statement

This work is part of the project call fishscan, where the main goal is design a system for remote monitoring of fishes size and weight distribution in aquaculture environments. during this project was develop a camera rig system consist in a underwater housing with a time of flight camera with LED light source and a 2D CCD grayscale camera as is shown in the fig 1.1.

as the main goal of the project is compute the biomass of the fish by computing the volume of it, taken the concept of mass density from the physics, using the relation between biomass and volume. Then, this problem of biomass estimation can be formulated as a problem of volume estimation of the fish. to achieve this objective, the first step is the fish detection in an 2D grayscale image. followed by a back-projection into the TOF image where the is possible to fit a 3D model to the detected fish. it is important to mention that

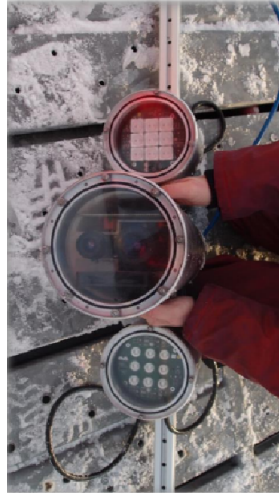


Figure 1.1.: Rig Camera System - TOF + CCD cameras

the approach assume in this work is due of highly noisy image acquire by the range imaging camera, which was adapted to work in a underwater environment, but as you can see in the 3D image shown in 1.2

although, the detection using the 2D intensity image alone cannot compute the volume of the fish because it is not depth invariant and the size is up-to-scale; This work will may use of the 2D intensity image as a First step, detecting the fish contour. The pipeline depicted in consist of three major steps that are: fish detection, contour extraction, volume-biomass estimation. In this project, we concentrate on the first step that is fish detection and contour extraction.

At this point we need to find a algorithm for fish detection and contour extraction that addresses the three major challenges of our problem, These are:

1. *Deformations*, THe algorithm must handle different motions of the which suggests that we are dealing with a deformable (articulated) object.
2. *Different Viewpoints*, As the fish move around its environment, the algorithm must be able to detect the fish from different perspectives.
3. *Occlusions*, The algorithm must also be able to handle occlusion, e.g. self occlusions, occlusion from another fishes and occlusion from object in the environment.

Based on the three challenges and the use of 2D intensity values obtained from the 2D CCD camera, we propose an approach inspired by two works from, the first, *Mixtures-of parts et al.* which describe a method for articulated human detection and human pose estimation in static images based on a representation of deformable part models. The main idea of their approach is to use a mixture of small, non-oriented parts, which describe a general, flexible mixture model that jointly captures spatial relations between parts locations and co-occurrences relations between part mixtures, augmenting standard pictorial structure models that encode just spatial relations. The second approach is propose in *Hinterstoisser et al. [2012]*, where they present a method for real-time 3D object instance de-

add im-
age ref ??

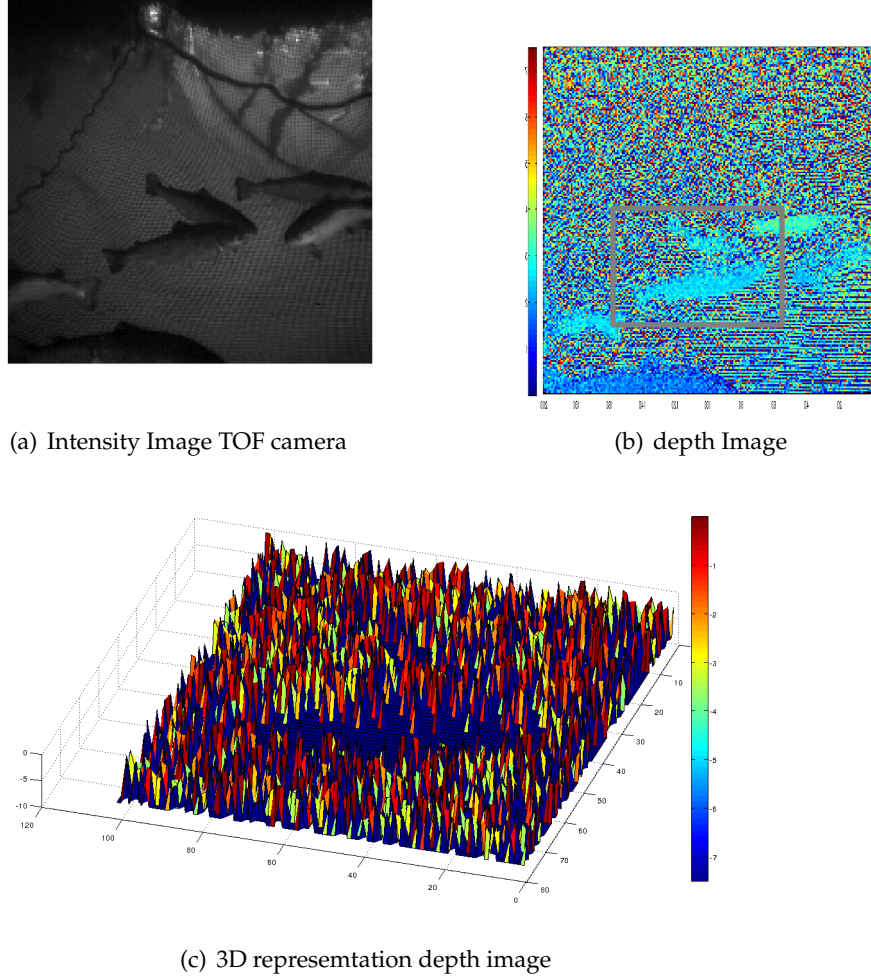


Figure 1.2.: TOF camera images.

tection that does not require a time consuming training stage, and can handle untextured objects. At its core, the approach presented is a novel image representation for template matching designed to be robust to small image transformations, This robustness is based on spread image gradient orientations and allows to test only a small subset of all possible pixel locations when parsing a image. In our approach, that can be considerer in the groups of machine learning method, the learning process process tries to understand the relation between the input X and output Y ; such that when the input X is given, it can predict the outcome Y . In our case, the input X are 2D grayscale intensity image containing fishes under different deformations and seen from different perspectives, while the expected output Y are the positions of the desired keypoints and corresponding contours for the detected fishes, However, in the learning stage, we require a large amount of data that shows this relations. We need a great amount of labeled 2D intensity image where the locations of the keypoints and contours are given. unfortunately, the current motion tracking systems from human pose estimation are not a valid solution to find ground truth data for fish

add short description of learning stage for part based and linemod

because the difference in behaviour as well as the difference in environment. Therefore, we address this problem by hand labeling real 2D deformed fishes images of a group of in a real scenario, where the fish are observed from different perspectives.

We evaluate the proposed method by matching the

explain method the evaluate result

Therefore, this thesis accomplished the following: create a labeled real fish images datasets that comprise of intensity images acquire by 2D CCD camera, learn esqueleton and contours from labeled datasets, predict keypoints and contours an unlabeled 2D fish images and verify the validity of the predictions that will be discussed in Chapter 3 . A overview of the related works is in Chapter 2 while the implementations details are presented in Chapter 4. The result is discussed in Chapter 5 and finally, we conclude in Chapter 6.

add ref

add ref

Continue from here

2. Related Work

2.1. Related Work

This thesis focuses on detecting fishes; therefore, this chapter concentrates on works that deal with object detection, specifically in underwater application. In [Shortis et al. \[2013\]](#) reviewed the state of arts on the field of identification and measurement of fish, using underwater stereo-video sequences. The most interesting approaches describes in this paper, are related with automated measurement of fish in video sequences as describe by [Tillett et al. \[2000\]](#) in what is one of the first published reports on successful, operational, automated measurement system. The technique is based on 3D Point Distribution (PDM), which are composed of landmark locations on the outline of the fish, in this case Atlantic Salmon held in a small aquaculture tank. The PDM specific to the species is developed from a small sample of fish defined by manual measurement of stereo-images, leading to a mean shape and an estimate of the variation based on principal component analysis. The PDM is independent of the scale and orientation, but is limited only to the silhouette of the fish and does not model the full body shape. This work, also analyse the methodologies for detection of fishes which comprises two steps: Identification and subsequent delineation of the fish outline. Most of the existing work on fish detection employ either the differences between successive images [Spampinato and Chen-Burger \[2008\]](#); [Tillett et al. \[2000\]](#) or histogram-thresholds to segment a varying number of candidate regions in the frames. Active contours(also called snake) are especially useful for delineating objects like fish bodies that are difficult to model with rigid geometric primitives. Moreover, active contours can be independent from edge gradients with flexibility in initialisation [Chan and Vese \[2001\]](#). The area-based active contour model [Chan and Vese \[2001\]](#) is based on the on techniques of curve evolution and level sets. While parametric active contours cannot handle automatic change of topology, level sets allow for splitting and merging in a natural way and are thus more suited for detection of an unknown number of fish in a video image. In [Shortis et al. \[2013\]](#), they also tackle the different technique apply in measurement. Underwater stereo systems are widely used to capture video of swimming fish for subsequent measurement. the simplest form of measurement is the fish snout to tail length which can be calculated if these two points can be identified in the stereo pair of images. in our days this is done manually in most cases, with a favourable orientation of the fish to the cameras and a multiple measurements within in the sequences of frames do improve the precision of the measurements. One important point to be noted is that fish are deformable and the euclidean distance from snout to tail changes as the fish swims. Template matching is one of the primitive methods that can be employed to accurately locate fish snout and tail in video frames. First, individual templates (usually rectangular image regions) centered on the snout and tail mid-points are extracted from sample videos. Then an efficient template matching strategy is employed to locate these templates in target videos. A certain degree of robustness against illumination changes can be achieved

by using correlation between template and image regions of interest, instead of taking their absolute differences [Mahmood and Khan \[2012\]](#). However, template based methods fail in the presence of perspective or affine transformations, requiring either use of multiple templates that capture appearance variations from different viewing angles, or using more sophisticated matching techniques that are invariant to affine or perspective transformations. These enhancements also significantly increase computational complexity of the template matching step. A better way of locating snout and tail is to use Haar-like features in a boosted classifier setup [Viola and Jones \[2001\]](#) that has shown high object detection accuracy, besides being able to operate in real-time. The method is in wide use for face detection. To train the classifier, manually cropped images of the target object (snout or tail) are used so that the classifier can learn which features (among a set of possibly thousands of features) can locate the target with high accuracy. These features, once learned, are then used to construct the object classifier that can locate the presence of the object in cluttered scenes. Due to their high detection speed and ability to perform a scale-space search, Haar classifiers are a promising candidate for locating snout and tail of fish in underwater images. The results of independent detection of the snout and tail using Haar detectors can be further improved using relationships between the detected snouts and tails, for instance by constraining the search for tail detection based on the results of snout detection and vice versa.

add ramanan and linemod approach

Furthermore, since we are detecting fishes, it is safe to limit our scope to articulated object detection. This lead to the idea that our problem is similar to human pose estimation. and specifically to

2.2. Problem Statement

Part II.

Methods and Implementation

3. Methods

3.1. Theoretical Background

There is no need for a latex introduction since there is plenty of literature out there.

3.2. Notation and Symbols

This section introduces the common mathematical notations and symbols used in this chapter. Normal formatting such as a and A are used to indicate integers or real numbers while letters in scripts such as \mathcal{A} are reserved for sets. In addition, we use the symbols \mathbb{R} for real numbers. For 3D vectors, we use the uppercase bold letters \mathbf{A} while for 2D vectors, we use the lowercase bold letters \mathbf{a} . If the vector is homogeneous, it will be explicitly define; otherwise, the vector is assumed to be inhomogeneous. A vector from point A to B is presented as \overrightarrow{AB} . Furthermore, matrices use monospace font such as \mathbf{A} . If the dimension are specified such as $\mathbf{A}_{m \times n}$, m would indicate the number of rows while n indicates the numbers of columns. Regarding accents, a hat on a vector such as \hat{a} or \hat{A} indicates the normalized unit vector which means that $\hat{a} = \frac{a}{\|a\|}$ or $\hat{A} = \frac{A}{\|A\|}$. A tilde on a 3D vector indicates that the last coordinate is removed; thus, the vector $A = (x, y, z)^T$ have $\tilde{A} = (x, y)^T$ which is the projection of A in the xy -axis, while the vector with homogeneous coordinate $B = (x, y, z, 1)^T$ have $\hat{B} = (x, y, z)^T$ which is the inhomogeneous coordinate of B . Lastly, a dot on top of variable indicates the converged value of the variable after an algorithm is performed. For instances, after using mean shift on a_i , it converges to a value \dot{a}_i

3.3. Part Based Model

There is no need for a latex introduction since there is plenty of literature out there.

3.4. Linemod

There is no need for a latex introduction since there is plenty of literature out there.

4. Implementation

Here starts the thesis with an introduction. Please use nice latex and bibtex entries [Lamport \[1994\]](#). Do not spend time on formatting your thesis, but on its content.

4.1. Fish Model

There is no need for a latex introduction since there is plenty of literature out there.

4.2. Linemod

There is no need for a latex introduction since there is plenty of literature out there.

Part III.

Results and Conclusion

5. Results and Discussion

5.1. Results and Discussion

In this chapter, we evaluate the proposed algorithm using real data acquire using the camera rig system shown in ?? and compare result with different approaches

define
approaches

There is no need for a latex introduction since there is plenty of literature out there.

5.2. Conclusion

There is no need for a latex introduction since there is plenty of literature out there.

6. Conclusion

Here starts the thesis with an introduction. Please use nice latex and bibtex entries [Lamport \[1994\]](#). Do not spend time on formatting your thesis, but on its content.

6.1. Discussion

There is no need for a latex introduction since there is plenty of literature out there.

6.2. Conclusion

There is no need for a latex introduction since there is plenty of literature out there.

Appendix

A. Detailed Descriptions

Here come the details that are not supposed to be in the regular text.

Bibliography

- Frank Asche and Trond Bjørndal. *The Economics of Salmon Aquaculture, Second Edition*. Wiley-Blackwell, 2011.
- Tony F. Chan and Luminita A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.
- Stefan Hinterstoisser, Cedric Cagniart, Slobodan Ilic, Peter Sturm, Nassir Navab, Pascal Fua, and Vincent Lepetit. Gradient response maps for real-time detection of textureless objects. *IEEE transactions on pattern analysis and machine intelligence*, 34(5):876–88, May 2012. ISSN 1939-3539. doi: 10.1109/TPAMI.2011.206. URL <http://www.ncbi.nlm.nih.gov/pubmed/22442120>.
- Leslie Lamport. *LaTeX : A Documentation Preparation System User's Guide and Reference Manual*. Addison-Wesley Professional, 1994.
- THOMAS A. LARSEN and FRANK ASCHE. Contracts in the Salmon Aquaculture Industry: An Analysis of Norwegian Salmon Exports, 2011. ISSN 0738-1360.
- Yajie Liu, Jon Olaf Olaussen, and Anders Skonhøft. Wild and farmed salmon in Norway-A review. *Marine Policy*, 35(3):413–418, 2011.
- Arif Mahmood and Sohaib Khan. Correlation-coefficient-based fast template matching through partial Elimination. *IEEE Transactions on Image Processing*, 21(4):2099–2108, 2012.
- Flexible Mixtures-of parts, Yi Yang, and Deva Ramanan. Articulated Human Detection with. pages 1–15.
- Mark R Shortis, Mehdi Ravanbaksch, Faisal Shafait, Philip Culverhouse, Danelle Cline, and Duane Edgington. A review of techniques for the identification and measurement of fish in underwater stereo-video image sequences. *SPIE Optical ...*, 2013. URL <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1691542>.
- Concetto Spampinato and YH Chen-Burger. Detecting, Tracking and Counting Fish in Low Quality Unconstrained Underwater Videos. *VISAPP (2)*, 2008. URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.63.7796&rep=rep1&type=pdf>.
- Robin Tillett, Nigel Mcfarlane, and Jeff Lines. Estimating Dimensions of Free-Swimming Fish Using 3D Point Distribution Models. *Computer Vision and Image Understanding*, 79: 123–141, 2000. ISSN 10773142. doi: 10.1006/cviu.2000.0847.

- P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 1, 2001.