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 enterely the result of my own work except where otherwise indicated. I have only used the resources given in the list of references.

München, den 4. Juli 2014 Andrés Sánchez

Acknowledgments

I have a great deal of people to thank for the presented work, but the most important one is my wife, who left everything behind in our home country to be with me in this adventure, Without her company, support and affection, I could not fulfill my dream of studying in Germany.

I thank Dr. Slobodan Ilic and Prof. Dr. Nassir Navab as they play important roles in the completion of this thesis. In addition, I acknowledge the help from the Computer Vision group at the Chair for Computer Aided Medical Procedures and Augmented Reality. Furthermore, I would like to specifically mention the contribution of David Tan, for sharing his knowledge during our project in the field of deformable object detection in underwater applications.

It is a pleasure to thank who made this possible. I want to thank the Chilean National Commission for Science and Technology ($CONICYT$), whose funding and confidence have done possible to carry out my project in Germany.

Also, I would like to thank to three compatriot, how make my stay in the TU München possible, Dr. Victor Castañeda for help me during my application to the program and the beginning of this process, Jose Gardiazabal for offering always a hand when I needed it and Eduardo Morral for offered me accommodation during the tough task of find a place to live in München. Finally, I would like to thank my family for their outstanding support for the entire duration of my course.

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 constinuously 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 ara arranged pre-harves, 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, compose of time of flight range camera and CCD grayscale camera, that film fishes in the cage for sertain 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 occlussion 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 grayscales images. Inspired by recent works in we develop a similar approach for fish detection. we use a

We evaluate the proposes method by computing difference between tha 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

Outline of the Thesis

Part I: Overview

CHAPTER 1: INTRODUCTION

This chapter presents an overview of the thesis and it 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: THEORICAL 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 Bj??rndal [2011], the production increased from 16 M in the 1970s to 142 M in 2008. In these statistical figures, the amount if 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 Bj??rndal, 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 Bj??rndal [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

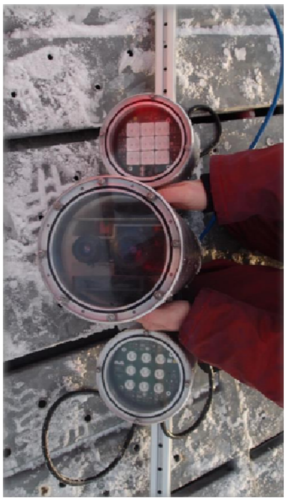


Figure 1.1.: Rig Camera System - TOF + CCD cameras

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.

add image ref **??**

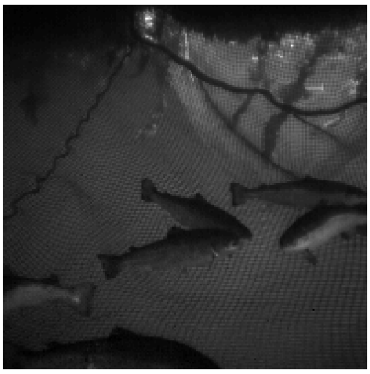
At this point we need to find a algorithm for fish detection and contour extraction that addresses the three major challenges of our problem, These are:

1. *Deformations*, THe algorithm must handle different motions of the which suggests that we are dealing with a deformable (articulated) object.

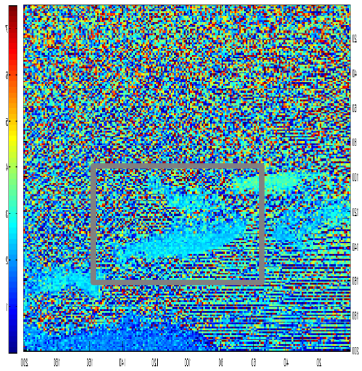
2. *Different Viewpoints*, As the fish move around its environment, the algorithm must be able to detect the fish from different perspectives.

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.

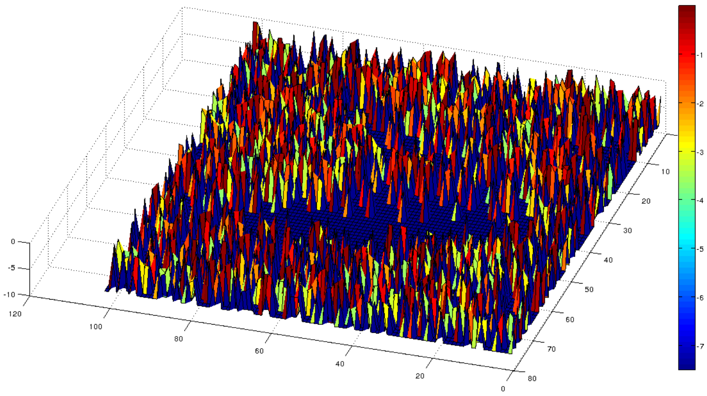
Based on the three challenges and the use of 2D intensity values obtained from the 2D CCD camera, we propose an approach inspired by two works from, the first, Mixtures-of parts et al. which describe a method for articulated human detection and human pose estimation in static images based on a representation 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-



(a) Intensity Image TOF camera



(b) depth Image



(c) 3D represemtation depth image

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

explain method the evaluate result

add ref

add ref

Continue from here

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

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 countours from labeled datasets, predict keypoints and contours an unlabeled 2D fish images and verify the validality 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.

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. Theorical Background

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

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 $\mathcal{A}$ 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 if 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 $\mathtt{A}$. If the dimension are specified such as $\mathtt{A}_{mxn}$, m would indicate the number of rows while n indicates the numbers of columns. Regardings 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 $\hat{A} = (x, y)^T$ 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. Tracking Workflow and Theory

The aim of this chapter is to define a basic workflow for the new tracking approach and provide the reader with relevant literature review. The workflow will be divided in to different functional processes. The requirements (Input) and the outcome (Output) of each process will be defined. Based on these requirements relevant theorical concepts will be discussed. The processes directly related to the problem statement will be discussed in detail, and a suitable approach will be suggested. The reader should note that this chapter will provide only the overview of the theorical concepts, implementation specific details will be covered in Part of the thesis. some where

3.3.1. Basic Workflow

3.1 shows the basic workflow of the detection process. The figure shows that once image is captured, it is supplied to the Part Based model detection process to extract 2D bounding

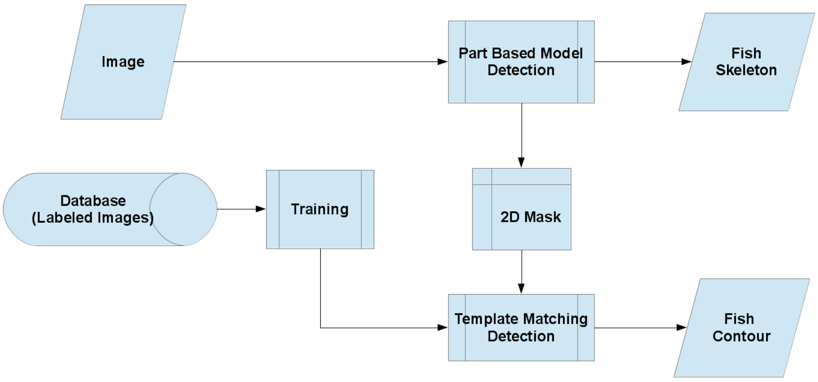


Figure 3.1.: A simplified workflow of the Detection process

boxes from the define part based fish model. Also from this process can be extracted the skeleton of the fish, After 2D mask are generated, template detection matching process will take the 2D masked image and in those defined areas will match the learned templates using the linemod algorithm propose in Hinterstoisser et al. [2012]. It is assumed that two detection process has been trained with the proper labeled fish images in a offline procedure. A short introduction to the terms, processes and their functionalities is given below.

- *Image*, this term refers to the two dimensional photographic image of the object captured by the camera. It is given as a input for the detection workflow. The reader should note that the image provided to the workflow come from the underwater rig system develop as part of the project. which suggests that we are dealing with a deformable (articulated) object.
- *Database*, This term represents all the preprocessed information that is readily available to the detection workflow. This information includes details regarding labeled data for training, trained model for Part Based detection process, and trained model for template matching process.

define

- *2D Mask*,
- *Part Based Model Detection*, this term refers to a broad class of detection algorithms used on images, in which various parts of the image are used separately in order to determine if and where an object of interest exists. Amongst these methods a very popular one is the constellation model which refers to those schemes which seek to detect a small number of features and their relative positions to then determine whether or not the object of interest is present.

define

- *Template Matching Detection*, This term refers to

define

- *Training*,

- *Fish skeleton*, define
- *Fish Contour*, define

3.3.2. Part Based model Detection

add images

add flowchart

redo the text to implement the problem

In this section will be discuss the method used to achieve the first task define in **??**, analysing our target object, the fish, which can be model like a deformable object with one main deformation axis, the one define by the backbone, Most fish move by alternately contracting paired sets of muscles on either side of the backbone. These contractions form S-shaped curves that move down the body. The pictorial structure framework arise as a feasible approach, which decomposes the appearance of the objects into local part templates, together with geometric constraints on pairs of parts, often visualized as a spring. when parts are parametrized by pixel location and orientation, the resulting structure can model articulation. This has been the dominant approach for human pose estimation. For our problem comparing with Full-body pose estimation where many degrees of freedom has to be estimated, The fish is a simplification due to the movement constraints, and also the absence of big limbs which are replace by small fins, those fins not vary greatly in appearance in compare with human limbs. in the broad class of part based detection algorithm, the one propose by Ramanan [2012] arise as the state of the arts, and is the one selected to be applied in our problem.

Model

adapt example

Let us write I for a image, $l_i = (x, y)$ for the pixel location of part i and t_i for the mixture component of part i. We write $i \in \{1 \dots K\}$, $l_i \in \{1 \dots L\}$ and $t_i \in \{1 \dots T\}$. we call t_i the "type" of part i. Our motivating examples of types include orientations of parts (e.g., a vertical versus horizontally oriented hand), but types may span out-of-plane rotations (front-view head versus side-view head) or even semantic classes (an open versus side-view head) or even semantic classes (an open versus closed hand). For notational convenience, we define lack of subscript to indicate a set spanned by that subscript (e,g,. $t = \{t1 \dots t_K\}$. For simplicity, we define our model at a fixed scale; at test time we detect people of different sizes by searching over an image pyramid.

Co-occurrence model: To score of a configuration of parts, we first define a compatibility function for part types that factors into a sum of local and pairwise scores:

$$S(t) = \sum_{i \in V} b_i^{t_i} + \sum_{ij \in E} b_{ij}^{t_i,t_j} \tag{3.1}$$

The parameter $b_i^{t_i}$ favors particular type assignments for part i, while the pairwise parameter $b_ij^{t_i,t_j}$ favors particular co-occurrences of part types. For example, if part types correspond to orientations and part i and j are on the same rigid limb, then $b_ij^{t_i,t_j}$ would favor consistent orientation assignments. Specifically, $b_ij^{t_i,t_j}$ should be a large positive number for consistent orientations t_i and t_j, and a large negative number for inconsistent orientations t_i and t_j.

Rigidity: we write $G = (V, E)$ for a (tree-structured) $K - node$ relational graph whose edges specify which pairs of parts are constrained to have consistent relations.Such a graph can still encode relations between distant parts through transitivity. For example, our model can force a collection of parts to share the same orientation, so long as the parts to share the same orientation, so long as the parts form a connected *subtree* of $G = (V, E)$. We use this property to model multiple parts on the torso. Since co-occurrence parameters are learned, our model learns which collections of parts should be rigid. We can now write the full score associated with a configuration of part types and positions:

$$S(t, l, t) = S(t) + \sum_{i \in V} \omega_i^{t_i} \cdot \phi(I, l_i) + \sum_{ij \in E} \omega_{ij}^{t_i, t_j} \cdot \psi(l_i - l_j) \tag{3.2}$$

where $\phi(I, l_i)$ is a feature vector(e.g, HOG descriptor **?**) extracted from pixel location l_i in image I. we write $\psi(l_i - l_j) = [dx\ dx^2\ dy\ dy^2]^T$, where $dx = x_i - x_j$ and $dy = y_i - y_j$, the relative location is defined with respect to the pixel grid and not the orientation of part i (as in classic articulated pictorial structures).

Appearance model: The first sum in 3.2 is an appearance model that computes the local score of placing a template $\omega_i^{t_i}$ for part i, tuned for type t_i, at location l_i.

Deformation model: The Second term can be interpreted as a šwitchingšpring model that controls the relative placement of part i and part j by switching between a collection of springs. Each spring is tailored for a particular pair of types (t_i, t_j), and is parametrized by its rest location and rigidity, which are encoded by $\omega_{ij}^{t_i, t_j}$.Our switching spring model encodes the dependence of local appearance on geometry, since different pairs of local mixtures are constrained to use different springs. Together with the co-occurrence term, it specifies an image-independent "prior" over part locations and types.

Inference

Inference corresponds to maximizing $S(I, l, t)$ from 3.2 over l and t. When the relational graph $G = (V, E)$ is a tree, this can be done efficiently with dynamic programming. To illustrate inference, let us re-write 3.2 by defining $z_i = (l_i, t_i)$ to denote both the discrete pixel location and discrete mixture type of part i:

$$\begin{aligned} S(I, z) &= \sum_{i \in V} \phi_i(I, z_i) + \sum_{ij \in E} \psi_i j(z_i, z_j), \\ where \quad \phi_i(I, z_i) &= \omega_i^{t_i} \cdot \phi(I, l_i) + b_i^{t_i} \\ \psi_{ij}(z_i, z_j) &= \omega_{ij}^{t_i, t_j} \cdot \psi(l_i - l_j) + b_{ij}^{t_i, t_j} \end{aligned} \tag{3.3}$$

From this perspective, it is clear that our final model is a discrete, pairwise Markov random field (MRF). When $G = (V, E)$ is tree-structure, one can compute $max_z\, S(I, z)$ with dynamic programming.

To be precise, we iterate over all parts starting from the leaves and moving "upstream" to the root part. We define kids(i) be the set of children of part i, which is the empty set for leaf parts. We compute the message part i passes to its parent j by the following:

$$score_i(z_i) = \phi_i(I, z_i) + \sum_{k \in kids(i)} m_k(z_i) \tag{3.4}$$

$$m_i(z_j) = \max_{z_i} \left[score_i(z_i) + \psi_{ij}(z_i, z_j)\right] \tag{3.5}$$

3.4 computes the local score of part i, at all pixel locations l_i and for all possible types t_i, by collectiong messages from the children of i. 3.5 computes for every location and possible type of part i, the best scoring location and type of its child part i. Once messages are passed to the root part $(i = 1)$, $score_1(z_1)$ represents the best scoring configuration for each root position and type. One can use these root scores to generate multiple detections in image I by thresholding them and applying non-maximum suppresion (NMS). by keeping track of the $argmax$ indices, one can backtrack to find the location and type of each part in each maximal configuration. To find multiple detections anchored at the same root, one can use $N - best$ extensions of dynamic programming. *Computation:* The computationally taxing portion od fynamic programming is 3.5. We rewirte this step in detail:

$$m_i(t_i, l_j) = \max_{t_i} \left[b_{ij}^{t_i,t_j} + \max_{l_i} score_i(t_i, l_i) + \omega_{ij}^{t_i,t_j} \cdot \psi(l_i, l_j)\right] \tag{3.6}$$

onee has to loop over LxT possible parent locations and types, and compute a max over LxT possible child locations and types, making the computation $O(L^2T^2$ for each part. When $psi(l_i - l_j)$ is a quadratic function (as is the case for us), the inner maximamization in 3.6 can be efficiently computed for each combination of t_i and t_j in $O(L)$ with a max-convolution or distance transform **?**. Since one has to perform T^2 distance transforms, message passing reduces to $O(LT^2)$ per part.

Learning

We assume a supervised learning paradigm. Given labeled positive examples I_n, l_n, t_n and negative examples I_n, we will define a structured prediction objective function similar. To do so, let us write $l_n = (l_n, t_n)$ and note that the scoring function 3.2 is linear in model parameters $\beta = (\omega, b)$, and so can be written as $S(I, z) = \beta \cdot \Phi(I, z)$. We would learn a model of the form:

$$\begin{aligned} arg \quad mim_{\omega,\xi_n \geq 0} \quad & \frac{1}{2}\beta \cdot \beta + C \sum_n \xi_n \\ s.t. \forall n \in pos \quad & \beta \cdot \Phi(I_n, z_n) \geq 1 - \xi_n \\ \forall n \in neg, \forall z \quad & \beta \cdot \Phi(I_n, z) \leq -1 + \xi_n \end{aligned} \tag{3.7}$$

The above constraint states that positive examples should score better than 1 (the margin), while negative examples, for all configurations of part positions and types, should score less than -1. The objective function penalizes violations of these constraints using slack variables ξ_n.

3.3.3. Template based Model

4. Implementation

Here starts the thesis with an introduction. Please use nice latex and bibtex entries Lamport [1994]. Do not spend time on formating 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

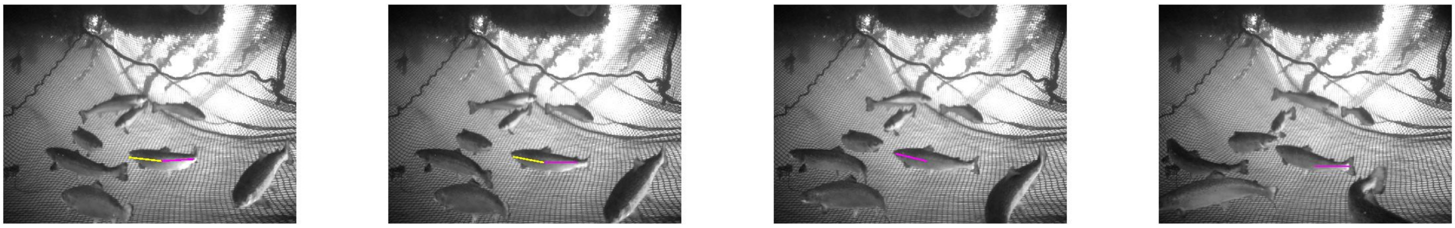
In this chapter, we evaluate the proposed algorithm using real data acquire using the camera rig system shown in **??** and compare result with different approaches

define approaches

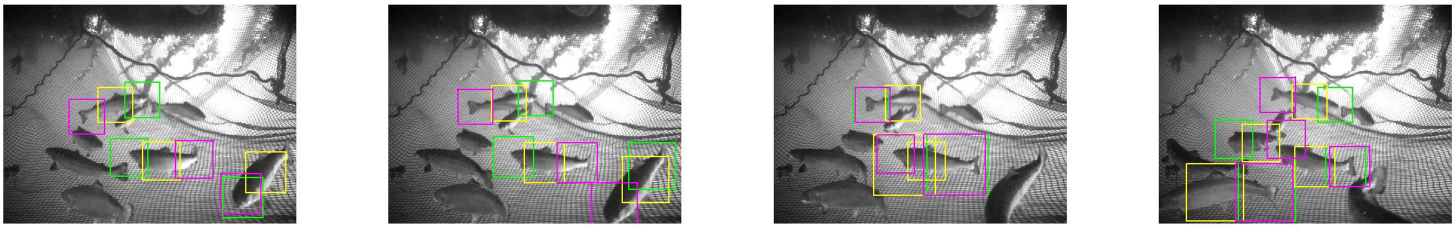
There is no need for a latex introduction since there is plenty of literature out there.

5.2. Conclusion

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

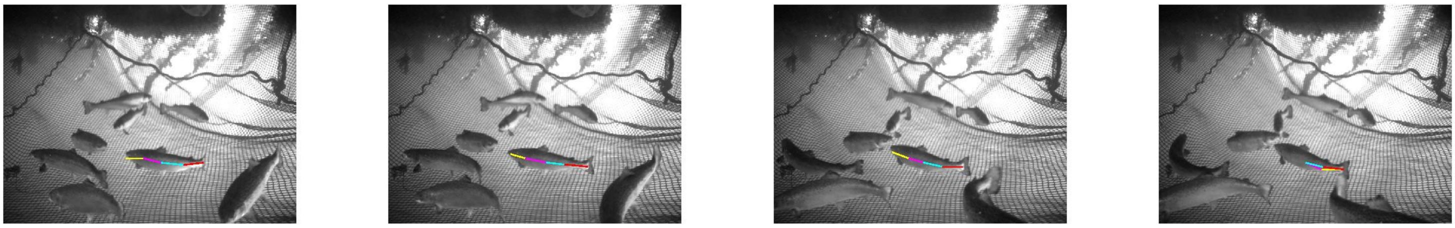


(a) Figure A (b) Figure B (c) Figure C (d) Figure D

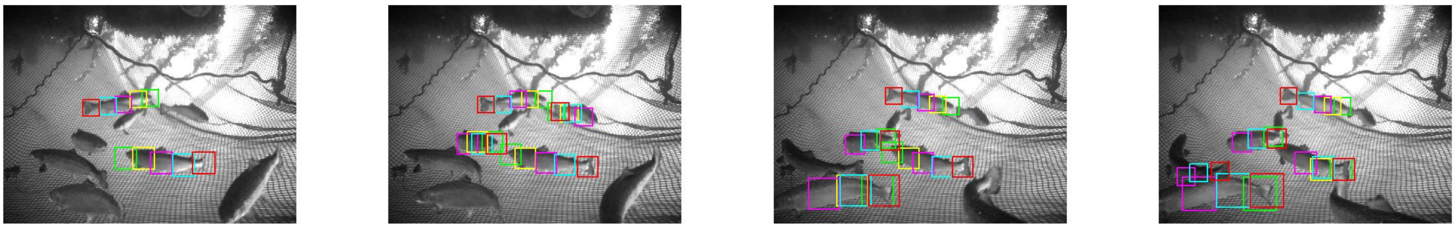


(e) Figure A (f) Figure B (g) Figure C (h) Figure D

Figure 5.1.: my caption



(a) Figure A (b) Figure B (c) Figure C (d) Figure D



(e) Figure A (f) Figure B (g) Figure C (h) Figure D

Figure 5.2.: my caption

6. Conclusion

Here starts the thesis with an introduction. Please use nice latex and bibtex entries Lamport [1994]. Do not spend time on formating 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 Bj??rndal. *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 Skonhoft. 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.

D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2879–2886, June 2012. doi: 10.1109/CVPR.2012.6248014. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6248014.

Mark R Shortis, Mehdi Ravanbakskh, 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.