

Master's Thesis

Estimating 3D Shape of the Head Skeleton of Basking Sharks Using Annotated Landmarks on a 2D Image

Martha Paskin

December 9, 2021

Dr. Daniel Baum (Zuse Institute Berlin)

Dr. Christoph von Tycowicz (Zuse Institute Berlin)

Prof. Dr. Konrad Polthier (Freie Universität Berlin)

Abstract

Basking sharks are thought to be one of the most efficient filter-feeding fish in terms of the throughput of water filtered through their gills. Details about the underlying morphology of their branchial region have not been studied due to various challenges in acquiring real-world data. The present thesis aims to facilitate this, by developing a mathematical shape model which constructs the 3D structure of the head skeleton of a basking shark using annotated landmarks on a single 2D image. This is an ill-posed problem as estimating the depth of a 3D object from a single 2D view is, in general, not possible. To reduce this ambiguity, we create a set of pre-defined training shapes in 3D from CT scans of basking sharks. First, the damaged structures of the sharks in the scans are corrected via solving a set of optimization problems, before using them as accurate 3D representations of the object. Then, two approaches are employed for the 2D-to-3D shape fitting problem—an Active Shape Model approach and a Kendall’s Shape Space approach. The former represents a shape as a point on a high-dimensional Euclidean space, whereas the latter represents a shape as an equivalence class of points in this Euclidean space. Kendall’s shape space approach is a novel technique that has not yet been applied in this context, and a comprehensive comparison of the two approaches suggests this approach to be superior for the problem at hand. This can be credited to an improved interpolation of the training shapes.

Kurzzusammenfassung

Riesenhaie zählen zu den effizientesten Filtrierern hinsichtlich des durch die Kiemen gefilterten Wasservolumens. Die Kiemenregion dieser Tiere besitzt eine markante Morphologie, die jedoch bisher nicht umfassend erforscht werden konnte, da es schwierig ist, reale Daten dieser Tiere zu erheben. Die vorliegende Arbeit zielt darauf ab, dies durch die Entwicklung eines mathematischen Formmodels zu ermöglichen, das es erlaubt, die 3D-Struktur des Schädelskeletts anhand von Landmarken, die auf einem 2D-Bild platziert werden, zu rekonstruieren. Die hierzu benötigte Tiefenbestimmung der Landmarken aus einer 2D-Projektion ist ein unterbestimmtes Problem. Wir lösen dies durch die Hinzunahme von Trainingsformen, welche wir aus CT-Scans von Riesenhaien gewinnen. Der Zustand der tomografierten Exemplare erfordert jedoch einen vorhergehenden Korrekturschritt, den wir mit Hilfe eines Optimierungsansatzes lösen, bevor die extrahierten Strukturen als 3D-Trainingsformen dienen können. Um die 3D-Struktur des Schädelskeletts aus 2D-Landmarken zu rekonstruieren, vergleichen wir zwei Ansätze – den sogenannten Active-Shape-Model (ASM)-Ansatz und einen Ansatz basierend auf Kendalls Formenraum. Während eine Form des ASM-Ansatzes durch einen Punkt in einem hochdimensionalen Euklidischen Raum repräsentiert ist, repräsentiert eine Form im Kendall-Formenraum eine Äquivalenzklasse von Punkten des Euklidischen Raumes. Die Anwendung des Kendall-Formenraumes für das beschriebene Problem ist neu und ein umfassender Vergleich der Methoden hat ergeben, dass dieser Ansatz für die spezielle Anwendung zu besseren Ergebnissen führt. Wir führen dies auf die überlegene Interpolation der Trainingsformen in diesem Raum zurück.

Acknowledgements

This master's thesis project was performed in collaboration with the Human Frontier Science Program (HFSP) project, titled "Integrating Materials, Behaviour, Robotics and Architecture in Giant Filter-Feeding Sharks".

First and foremost, I would like to express my utmost gratitude to my supervisors at the Zuse Institute Berlin, Dr. Daniel Baum and Dr. Christoph von Tycowicz, who helped me immensely during my graduate research through their invaluable expertise and constructive feedback. Daniel's dedicated involvement, creative insights and empathetic approach, enabled me to realize my full potential. Christoph raised several precious points throughout our discussions, which contributed greatly to the scope of this thesis. I would like to thank Prof. Dr. Konrad Polthier, who supported my thesis proposal and offered his treasured advice. Also, my colleagues, Felix Herter and Justus Vogel, who helped me greatly with their valuable suggestions.

In addition, I would like to acknowledge my collaborators, Dr. Mason Dean (City University Hong Kong) and Dr. Sean Hanna and his team (University College London), who gave me the opportunity to gain further knowledge, and apply my skills in a multi-disciplinary framework. This inspired me creatively and broadened my perspective. I especially thank Dr. Mason Dean for proposing this research and offering his expertise in the field of biology.

Last but not least, I thank my family and friends, especially my sister, for their continuing support and trust in me, which helps me become a better person every day.

Declaration of Authorship

I hereby declare that I have written the present thesis independently, as a result of my research. All the used sources have been specifically acknowledged, and no other sources, other than those, have been used.

Date:

Signature

Contents

1	Introduction and Related Work	6
1.1	The Basking Shark—significance and morphology	7
1.2	Contributions	10
2	Data Preparation	12
2.1	Image Processing	12
2.1.1	Image Segmentation	12
2.1.2	Skeleton creation	14
2.2	Data Correction and Augmentation	14
2.2.1	Skeleton splitting	14
2.2.2	Skeleton correction	15
2.2.3	Data augmentation	21
3	Active Shape Model Approach	25
3.1	The Original Active Shape Model	25
3.1.1	Introduction	25
3.1.2	Method	25
3.2	Estimating 3D shape from 2D landmarks using ASM	28
3.2.1	Introduction	28
3.2.2	Problem Formulation	28
3.2.3	Non-convex Formulation	30
3.2.4	Convex formulation	30
3.3	Implementation	33
4	Kendall’s Shape Space Approach	35
4.1	Kendall’s Shape Space	35
4.1.1	The Fréchet Mean	37
4.2	Problem Formulation	38
4.3	Solving the Optimization Problem	38
4.4	Implementation	38
5	Results	40
5.1	Exactness of recovery	40
5.2	Robustness	41
5.3	Application on real-world data	46
6	Discussion and Outlook	51
6.1	Comparing the approaches	51
6.2	Basking shark head skeleton data	52
6.3	Outlook	52

1 Introduction and Related Work

In recent years, the problem of estimating the 3D shape of an object from a 2D view has gained great attention from the computer vision community. In this thesis, we aim to retrieve the 3D shape of a non-rigid object from a single monocular 2D image. This is a promising but challenging problem, with widespread applications. A few examples include autonomous driving, robot assisted surgery and navigation, and in computer aided design (CAD). Estimating the depth of a 3D object from a single 2D view is an ill-posed problem. Common approaches to deal with this include:

1. Using a series of 2D images [How04, CWLZ13] or multiple cameras [LMPF07, JLT⁺12, MKGH15]: This is a widely used approach, e.g. in photogrammetry, as it greatly reduces the depth ambiguity by providing different view-points of a 3D object, via a stream of 2D images or multiple calibrated cameras. A common practise is to extract silhouettes from multiple 2D images, and construct a 3D shape using volume intersection [BL01] or a voxel-based approach [CBK03]. Other approaches include identifying corresponding points in 2D images for calculating depth [CWLZ13, PF01, AACM14]. The reader may refer to the book by Theo Moons et al. [MVG09] for details about this approach.
2. Physical modelling of the object [SNP16, AB15]: The object is described as a kinematic tree consisting of segments connected by joints. Each joint contains some degrees of freedom (DOF), indicating the directions it is allowed to move. The object is modelled by the DOF of all joints. Lengths constraints can also be added, such as limb lengths constraints in a human pose. This helps reduce the depth ambiguity by constraining the relative position of the joints.
3. Shape spaces [ZLHD15, RKS12, MM06]: This approach estimates the 3D shape of an object by interpolating through a set of pre-defined 3D shapes including the possible deformations of the object. Given a 2D projection of the object, the aim is then to find the ideal camera parameters and the shape coefficients describing the interpolation.

In this thesis, we use the third approach. This is inspired by the visual memory of humans, which, paired with our binocular vision, helps in depth perception of 3D objects. The set of pre-defined 3D shapes of an object is used as a replacement of its visual memory to estimate its 3D shape from a single 2D view.

We aim to estimate the 3D shape of the head skeleton of a basking shark, the significance of which is provided in the next section. For this, we use two different approaches—the first inspired by the Active Shape Model (ASM) [CTCG95], and the second by Kendall’s Shape Space (KSS) [Ken84]. In both approaches, an object is described by a set of annotated landmarks in 3D, and a set of pre-defined shapes is used to form a “shape space”, used to estimate an unknown 3D shape of the object. The annotated landmarks usually mark the significant

parts of the object. The goal is to solve the 2D-to-3D fitting problem, that given a set of annotated landmarks on a 2D image of the object representing the projection of the unknown 3D shape on a plane, estimate the latter. In the case of the head skeleton of a basking shark, the landmarks represent the end points of its constituting parts. The motivation of our work is similar to the publication by Rygg et al. [RCA⁺13], where they use high-resolution micro-Computed Tomography (CT) and magnetic resonance imaging (MRI) scans of the head and olfactory chamber of hammerhead sharks, to study the hydrodynamics in their nasal region.

The ASM has been used for many 2D-to-3D shape fitting problems [ZLHD15, RKS12, BHB00, HR12] and is a reliable approach. The details of this are provided in Section 3. The KSS approach, however, is a promising novel one, and has not yet been used in this context. In the ASM approach, each shape is represented by a vector in the Euclidean space, and the unknown shape is estimated as a linear combination of the pre-defined shapes.

In the KSS approach, each shape is represented as a point on a Riemannian manifold, and the unknown 3D shape is estimated as a weighted intrinsic mean of the points on the manifold representing the pre-defined shapes. This offers a better interpolation compared to the ASM approach, and is not view-dependent. Section 4 explains the details. Kendall’s shape space has previously been applied mainly in the field of geometric morphometrics. Nava-Yazdani et al. [NYHSvT20, NYHvT19] perform statistical analysis of epidemiological data and study femoral longitudinal data using geodesics on Kendall’s shape space. Amor et al. [ASS15] represent human bodies by a dynamic skeleton, and study their movement through trajectories on Kendall’s shape space. In a similar context, Friji et al. [FDC⁺21] and Hosini et al. [HBA20] propose a geometric deep learning framework for analyzing the movement of 3D skeletons over time using trajectories on Kendall’s shape space.

We compare the two approaches with regard to exactness of recovery, robustness and performance on real data. Our focus is multi-disciplinary, with applications in bio-mechanics, architecture and robotics.

1.1 The Basking Shark—significance and morphology

Cetorhinus maximus, commonly known as basking shark, is the second largest shark in existence. Being a suspension or filter feeder, it feeds on microscopic animals called zooplankton, by filtering two million liters of water per hour through its gills. Unlike other filter feeding sharks, the basking shark is a *ram feeder*, which means that it move forward with its mouths wide open, engulfing the prey along with the water surrounding it. David W. Sims provides a review of its biology [Sim08]. What sets basking sharks apart is their filtering efficiency, which is predictably higher than that of other filter feeders. From its feeding pictures and videos, one can observe the unique “flaring” of the gills, which completely changes the shark’s head shape, expanding it greatly (see Figure 1).



(a)



(b)

Figure 1: (a) Front view of a basking shark (*Cetorhinus maximus*) in feeding position. (b) Side view of a basking shark in closed mouth (left) and open mouth (right) position (courtesy of Nicholas Payne, Trinity College Dublin)

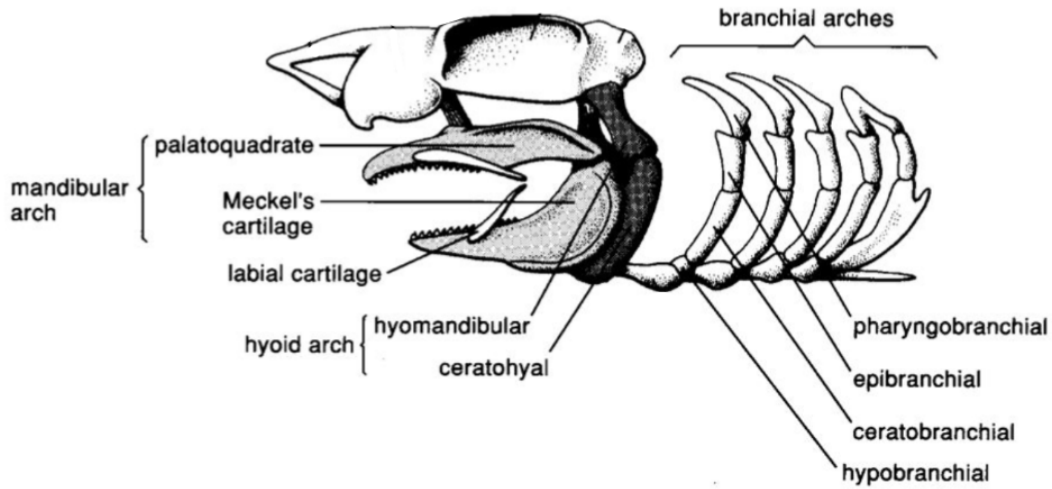


Figure 2: Labeled head skeleton of a shark.

Little is known about the performance of the filtering mechanism of basking sharks. They have historically been hunted for their meat and liver oil which has numerous industrial uses, and for their large fins which have a high demand for collectors and in shark-fin markets. This has caused their numbers to decline and the species is now listed as endangered. In many countries including the United States of America and the United Kingdom, basking sharks are heavily protected which makes injuring, harassing and killing them a punishable offence. Hence, the morphology of the basking shark has not been studied in great detail, to the best of our knowledge.

We were provided with real-world data of basking sharks in form of five CT scans, as well as videos and images in various poses and from different viewpoints. The goal was to use these data to build a 3D model of their branchial region, and use it to gain insights about the relative movement of the hyoid arch, the epibranchials, ceratobranchials and the pharyngobranchials (see Figure 2). This would help test the hypothesis about its filtering efficiency, and the findings could be used as bio-inspiration for large scale, potentially non-clogging dynamic filters. We do not include the mandibular arch for model simplicity, as its movement is largely similar to that of the hyoid arch. We also exclude the hypobranchials, due to their small size.

We have modelled the skeletal parts of interest as a piece-wise linear skeleton (Figure 3), composed of vertices (a set of nodes in 3D space) and edges (segments connecting the nodes). Each skeletal part is represented by an edge joining its end point points, and connected parts of the skeleton share a common vertex. The shape of each instance of our 3D object, the head skeleton of a basking shark, can therefore be represented by a set of vertex points in 3D. Section 2 explains the creation of some shapes of our 3D object, using CT scans of basking sharks. These shapes are later used in the 2D-to-3D fitting problem.

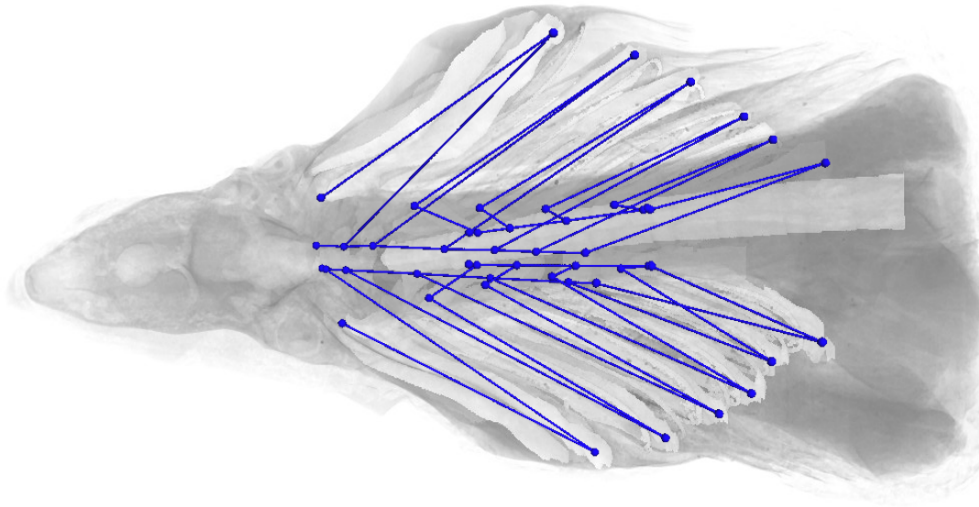


Figure 3: The piece-wise linear skeleton (blue) of a basking shark, overlaid on its CT scan.

We have used this piece-wise linear representation to move the different parts of the head skeleton via an optimization problem that tries to preserve the length of the parts. The details are explained in Section 2.

1.2 Contributions

The following contributions are made in this thesis:

- Kendall's shape space is applied for a 2D-to-3D fitting problem, which estimates the 3D shape of an object using annotated landmarks on a single 2D image. This is a novel approach to the problem, and produces promising results.
- The results obtained from Kendall's shape space approach are compared to the ones obtained using the active shape model, which is a frequently used approach for the 2D-to-3D shape fitting problem.
- A method to correct the damaged structures of the basking shark specimens in CT scans has been developed. This is done by representing the 3D volumetric skeletal parts of interest, in a CT scan, as a spatial graph consisting of vertices and edges. Optimization problems are then used to change the position of the vertices, while trying to preserve the lengths of the edges. Once corrected, the spatial graphs can be converted back to 3D volumetric data. This helps in creating an anatomically plausible model of

the head skeleton of a basking shark, which can be applied to real-world data in form of 2D images of undamaged basking sharks.

- The aforementioned optimization problems are used to open the mouth of the sharks in the CT scans. This is essential to the project and solves the problem of lack of real-world data in form of CT scans of basking sharks in open mouth position.

This thesis is structured as follows: Section 2 describes the creation of the pre-defined 3D shapes of the head skeleton of basking sharks, to be used in the 2D-to-3D fitting problem. Sections 3 and 4 explain the details of the ASM and KSS approaches, respectively. Section 5 described and compares the results obtained from the two approaches. Finally, Section 6 concludes this thesis with a discussion and outlook.

2 Data Preparation

This section describes the process of using real-world data in form of CT scans (represented as a 3D volumetric scalar field of size $\approx 512 \times 512 \times 1300$) to create a simplified representation of the parts of the head skeleton of a basking shark involved in its feeding motion, in form of a piece-wise linear skeleton. As mentioned in Section 1.1, the relative movement of different skeletal regions of basking sharks is not very well studied due to their endangered status and heavy protection. The only way to get hold of these creatures is the unfortunate event when one of these dead sharks is washed ashore. These specimens are, however, not in their original shape due to natural rotting and collisions with other objects. Moreover, fitting them into a CT scanner causes additional damage to their structure. Figures 4 and 5 show usable and unusable scans of basking sharks, respectively. Real world data in form of CT scans (provided by Mason Dean from City University Hong Kong), combined with high quality (monocular) image and video data, was used to create the 3D model. The regions of interest were isolated from the CT scans using data segmentation, and a geometric representation of these regions was created in form of spatial graphs consisting of nodes and segments. This process involved a number of steps, the details of which are presented in this section. This procedure was carried out using the data visualization software, Amira [SWH⁺05].

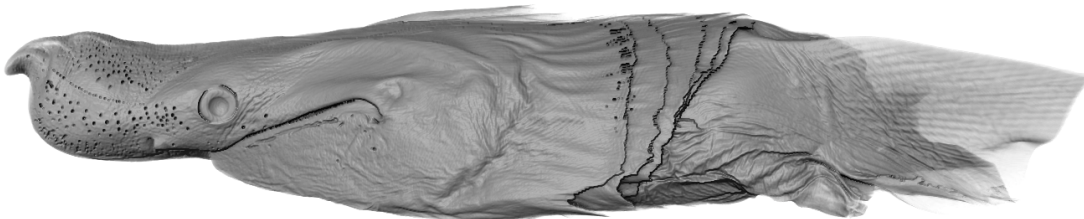
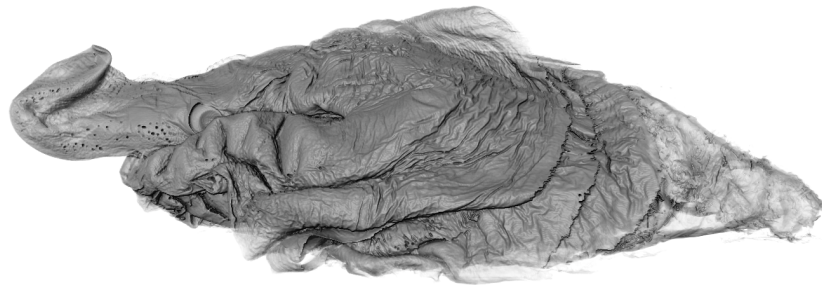


Figure 4: Example of a usable basking shark CT scan (side view)

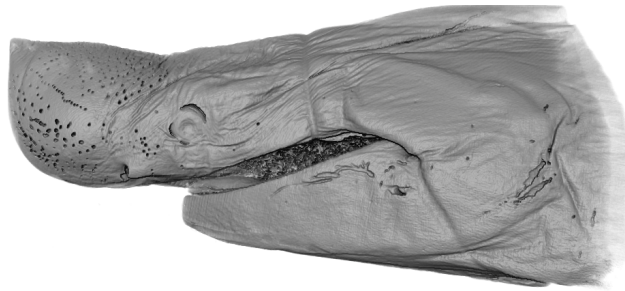
2.1 Image Processing

2.1.1 Image Segmentation

The first step in modelling the head skeleton of a basking shark was isolating it from the CT scans. For this, the CT scans were segmented using a mixture of region growing and manual segmentation with interpolation. The former was used to accelerate the process of voxel selection in a region's interior, and the latter was used to select the boundary of a region in a single cross-section. We could not rely solely on automatic segmentation methods, as the signal-to-noise ratio and the resolution of the scans was low (Figure 7), which caused the boundaries of regions to be blurred. Even the application of image filters did not solve this problem. Hence, at least the area close to the boundary of each region, had to be manually segmented. This was the most time consuming part of data extraction. Figure 6 shows the segmented regions of one of the CT scans.



(a)



(b)

Figure 5: Side views of unusable basking shark specimens:(a) damaged and (b) incomplete.

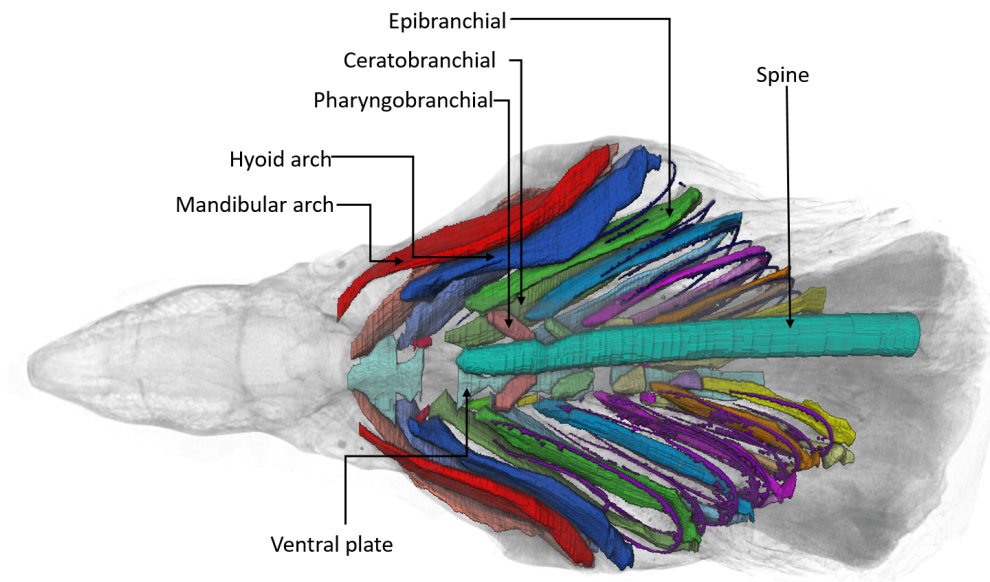


Figure 6: (Top view) Labeled segmented regions of the head skeleton of a basking shark (see Figure 2). The epibranchial, ceratobranchial and pharyngobranchial is labeled for the first gill arch only.

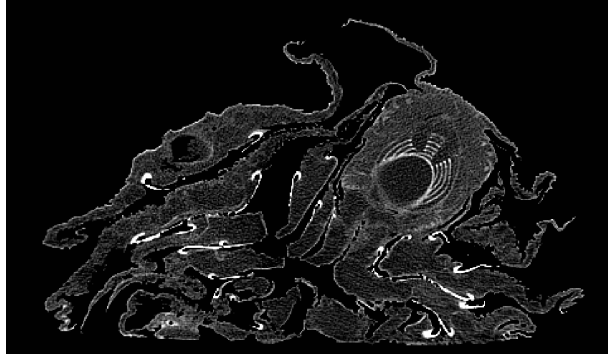


Figure 7: A cross-section of a noisy basking shark CT-scan.

The segmentation was performed using available information on the anatomy of other sharks and filter feeders [MPTS14, CHCF18], and was confirmed by Mason Dean.

2.1.2 Skeleton creation

Once the regions responsible for feeding were isolated using segmentation, these were ready to be studied further. To simplify the process, an abstract representation of this region was created in the form of a spatial graph consisting of nodes and segments. This led to the creation of a piece-wise linear skeleton of the shark (see Figure 8). The nodes of the skeleton were placed manually by selecting the end points of each region. The regions of the head skeleton which were not of interest to us were not incorporated in the skeletons. These include the mandibular arch, whose movement is largely similar to that of the hyoid arch and the spine, which remains stationary. This left us with the hyoid arch and the five gill arches (see Figure 2).

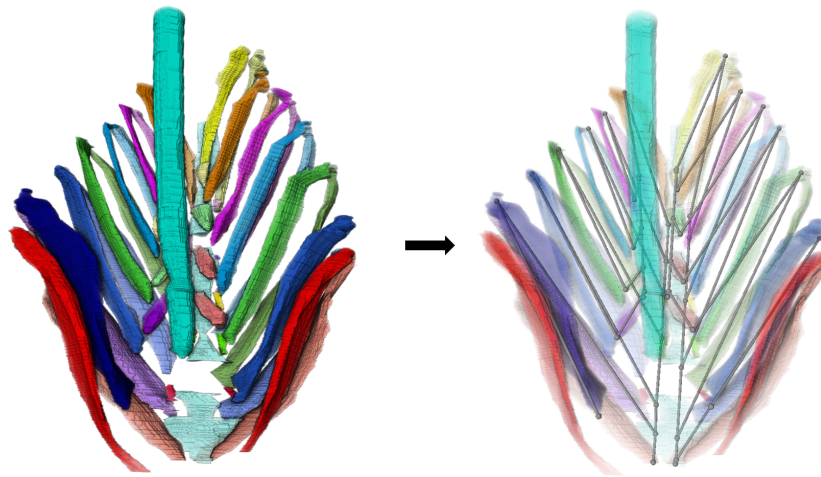
This simplified representation is advantageous, as it encompasses the vital properties of the original skeleton, which include the lengths of the segments and the connections between them. This information enabled us to study the relative movement of the skeletal segments. Moreover, each specimen could be represented uniquely by a set of 3D landmarks (nodes of the graph).

2.2 Data Correction and Augmentation

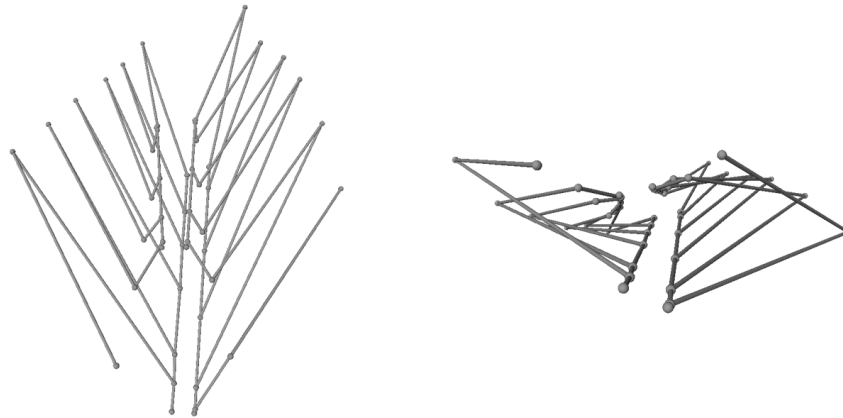
Initially, we were provided with five CT scans for this project but due to damaged and incomplete scans, only three of them could be used. Even the three usable scans had some amount of damage, and in order to create a plausible model, corrections to their structure were needed. We were able to do this by positioning the spine correctly using an optimization problem, which also helped in data augmentation. This section describes the procedure in detail.

2.2.1 Skeleton splitting

Like most fish, the basking shark is bi-laterally symmetrical, i.e., the right and the left halves are identical. Hence, the skeletons could be split into two halves, which



(a)



(b)

Figure 8: (a) Creating a piece-wise linear skeleton of a basking shark, using the segmented regions. (b) Top(left) and front(right) views of the skeleton.

could then be mirrored to create a full skeletons. This is a logical assumption and was confirmed by Mason Dean. Each half skeleton consists of 26 nodes and 29 segments (Figure 10).

2.2.2 Skeleton correction

The shark scans were not perfectly straight and the spine was slightly curved in most specimens. As the epibranchials connect to the spine via the pharyngobranchials and soft-tissue, they too were curved. This lack of bi-lateral symmetry can clearly be seen in Figure 8(b). Hence, to create a plausible 3D model of basking sharks, it was crucial to correct the spine positioning in the half skeletons.

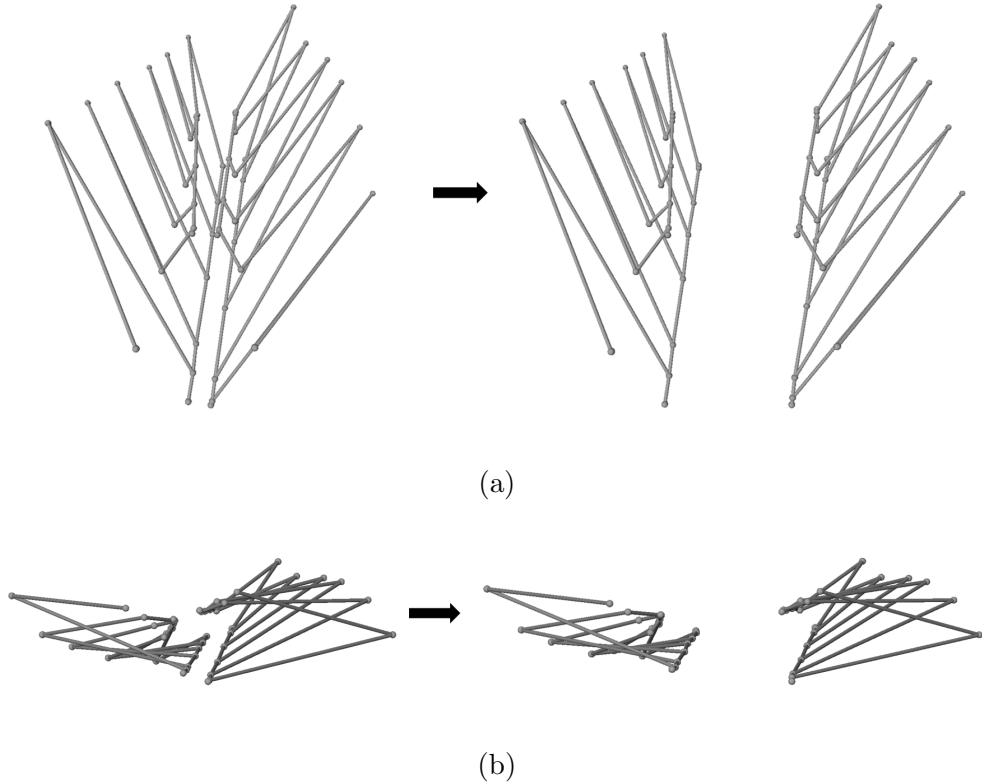


Figure 9: Splitting skeleton into right and left halves. (a) Top-view (b) Front-view.

Spine Correction

Note that the anatomical spine of the shark was not included in the skeleton for simplicity. We will, however, refer to the collection of segments in Figure 10 as “spine” for ease of notation, and will represent it by the line connecting the end nodes. It is intuitive that the movement of the rigid anatomical spine is identical to that of this artificial spine, as the latter is nothing but the former, excluding the soft-tissue which connects it to the pharyngobranchials.

The segments that constitute the spine have fixed length, and represent the length between the nodes of the pharyngobranchials that connect to the spine. Moreover, being rigid structures, the skeletal components (epibranchials, ceratobranchials, pharyngobranchials, and ventral plate) have fixed length. This gave rise to the idea of artificially moving the spine into a “correct” position, along with all skeletal components connected to it. For this, we used optimization problems which try to preserve the lengths of all the skeletal segments, while moving the nodes to a new position. This procedure is described in Algorithm 1, where we use the node order described in Figure 10(a), and is visualized in Figure 11. Due to steps 4 and 5 in Algorithm 1, a total of ten optimizations need to be performed to determine C_i and P_j , $\forall i \in \{h, 1, 2, 3, 4, 5\}$ and $j \in \{1, 2, 3, 4\}$. Table 1 describes the errors obtained for each measurement, for moving the spine in Figure 11.

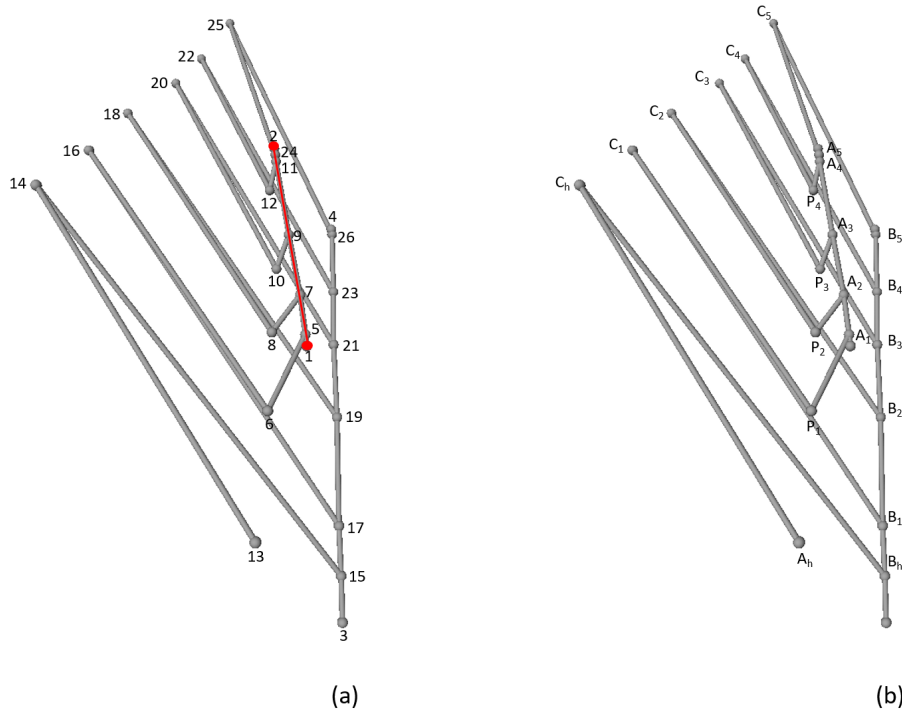


Figure 10: (a) The segment in red is referred to as the “spine” , and is represented by the end nodes (1 and 2). This segment consists of smaller segments, connecting the intermediate nodes. (b) A half-skeleton with the sets A , B , C and P labeled.

Algorithm 1 Move spine of half skeleton

- 1: Manually select two points, s_1 and s_2 , for the new position of the spine. Node 1 is placed at point s_1 , and the vector $s_2 - s_1$ gives the direction of the new spine. Node 2 is placed at the point $s_1 + l_s * (s_2 - s_1)$, where l_s is the length of the original spine and is preserved.
 - 2: The intermediate nodes, 5, 7, 9, 11 and 24, are placed on the new spine, preserving the length between them.
 - 3: The transformation that was applied to node 1 to move it to the new position, is applied to node 13.
 - 4: The nodes 14, 16, 18, 20, 22 and 25, representing the hyoid/gill arch mid points, are obtained by solving for C_i^{new} , $i \in \{h, 1, 2, 3, 4, 5\}$, in Algorithm 2.
 - 5: The nodes 6, 8, 10 and 12, connecting the pharyngobranchials to the epi-branchials, are obtained by solving for P_j^{new} , $j \in \{1, 2, 3, 4\}$, in Algorithm 3.
-

Algorithm 2 Optimizations for C^{new}

Require: Sets (see Figure 10(b)) :

$$\begin{aligned} A &= \{A_h, A_1, A_2, A_3, A_4, A_5\}, A_i \in \mathbb{R}^3, \\ B &= \{B_h, B_1, B_2, B_3, B_4, B_5\}, B_i \in \mathbb{R}^3, \\ a &= \{a_h, a_1, a_2, a_3, a_4, a_5\}, a_i = d(C_i, A_i) \in \mathbb{R}, \\ b &= \{b_h, b_1, b_2, b_3, b_4, b_5\}, b_i = d(C_i, B_i) \in \mathbb{R}. \end{aligned}$$

for $i \in \{h, 1, 2, 3, 4, 5\}$ **do**

$$C_i^{new} = \arg \min_{c \in \mathbb{R}^3} (d(c, A_i) - a_i)^2 + (d(c, B_i) - b_i)^2 \quad (1)$$

end for

Algorithm 3 Optimizations for P^{new}

Require: Sets (see Figure 10(b)) :

$$\begin{aligned} A &= \{A_1, A_2, A_3, A_4\}, A_i \in \mathbb{R}^3, \\ C &= \{C_1, C_2, C_3, C_4\}, C_i \in \mathbb{R}^3, \\ e &= \{e_1, e_2, e_3, e_4\}, e_i = d(C_i, P_i) \in \mathbb{R}, \\ f &= \{f_1, f_2, f_3, f_4\}, f_i = d(P_i, A_i) \in \mathbb{R} \end{aligned}$$

for $j \in \{1, 2, 3, 4\}$ **do**

$$P_j^{new} = \arg \min_{p \in \mathbb{R}^3} ((d(p, C_j) - e)^2 + (d(p, A_j) - f)^2) \quad (2)$$

end for

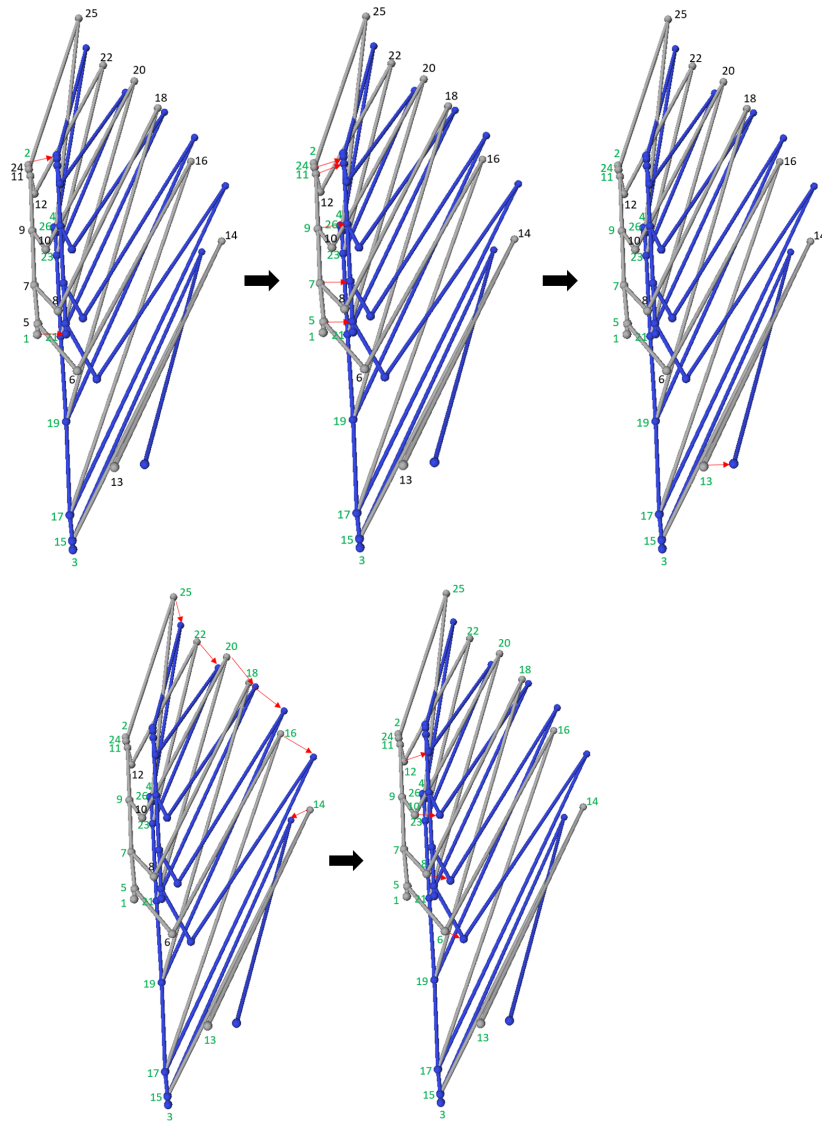


Figure 11: Moving the spine to a new position, following steps 1-5 (visualized from top left to bottom right). The gray and the blue skeletons represent the original and the new (moved spine) skeletons, respectively, the green nodes represent the nodes which are moved in each step and the red arrows represent the movement of the nodes.

Table 1: Optimization Errors for Moving Spine in Figure 11

Point	Error in Optimization
C_h	3e-6
C_1	3e-6
C_2	3e-6
C_3	3e-6
C_4	3e-6
C_5	3e-6
P_1	2e-6
P_2	2e-6
P_3	2e-6
P_4	1e-6

Iterations: 409 ; Computation time: 0.02 seconds.

The lower left and upper right coordinates of the 3D bounding box of the original half skeleton in Figure 11 are $(-21.39, -151.70, -994.86)$ and $(169.16, -62.71, -528.15)$, respectively.

Note that the nodes on the ventral plate, 3, 4, 15, 17, 19, 21, 23 and 26, remain the same when moving the spine. As the sets b , e and f refer to the lengths of the ceratobranchials, the pharyngobranchials and the epibranchials respectively, they must be fixed during the optimizations (1) and (2). We also fixed a in order to restrict the solution space. The optimization problems are sensitive to initialization as there does not exist a unique solution. Hence, the sets of points C in Algorithm 2 and P in Algorithm 3 were initialized to their positions in the original skeleton in order to obtain a solution close to the original position.

In addition to moving the spine of the skeleton, the above method can be modified to move the ventral plate, by placing its end nodes, 3 and 4, to the desired position. The intermediate nodes on the ventral plate were placed accordingly and the sets of points, C and P , are obtained by solving the optimization problems described in steps 4 and 5.

Virtually opening the Mouth

The above method moves the end nodes of the spine to an anatomically legitimate position to correct the damaged skeletons. The same method can be used to position the spine higher, thereby opening the mouth of the shark. Similar to above, the lengths of anatomical segments are preserved with no extra constraints. Figure 12 shows the obtained result and Table 2 describes the optimization errors. This computation was done in an early stage of this thesis, when even less was known about the movement of the the gill arches and when placing the spine higher, it was placed parallel to the ventral plate. Upon further investigation during the course of this thesis, it was revealed that this is an incorrect motion. This is the reason for the large error when optimizing for C_5 in Table 2. As this does not interfere with the goal of this thesis, this error is ignored.

Moving the spine and the ventral plate was performed by manually selecting the end nodes, and since the optimizations in steps 4 and 5 are sensitive to initialization, the new skeletons were not perfect. For example, in the open mouth position, the gill arches were sometimes folded backwards, which is an inaccurate configuration. In order to fix this issue, the option of rotating the gill arches, hyoid arch and pharyngobranchials was incorporated. This was done using the Rodriguez rotation formula [Rod40]. It states that given a vector $v \in \mathbb{R}^3$ and the axis of rotation described by the vector $e \in \mathbb{R}^3$, the vector $v_{rot} \in \mathbb{R}^3$, which is the vector v rotated about e by an angle of θ , is computed as:

$$v_{rot} = v \cos \theta + (\hat{e} \times v) \sin \theta + \hat{e}(\hat{e} \cdot v)(1 - \cos \theta),$$

where \hat{e} is the unit vector in the direction of e . Table 3 describes the vectors v and e for rotating the gill arches, hyoid arch and the pharyngobranchials. For different values of θ , the rotation changes the positions of the sets of points, C and P , while preserving the lengths of the skeletal segments. With this, we now had the tools to create numerous skeletal configurations by moving the spine, the ventral plate and also rotating the gills and the pharyngobranchials.

2.2.3 Data augmentation

The ability to correct the skeletons by adjusting the spine nodes as well as rotating certain skeletal regions, allowed for generating new skeletal configurations. These were then mirrored to create full skeletons. A nine fold increase in the number of configurations (from 3 to 27) was achieved using this approach, including the attainment of skeletal configurations in open mouth positions. Hence, along with solving the issue of very limited data, this also helped to correct the damaged skeletons and artificially open the mouth of the shark, which is a movement essential for this project. Figure 14 visualizes some of the final skeletons in various configurations. It should be noted that many more configurations can be created using the aforementioned methods. However, for the scope of this thesis, 27 configurations were considered to be sufficient.

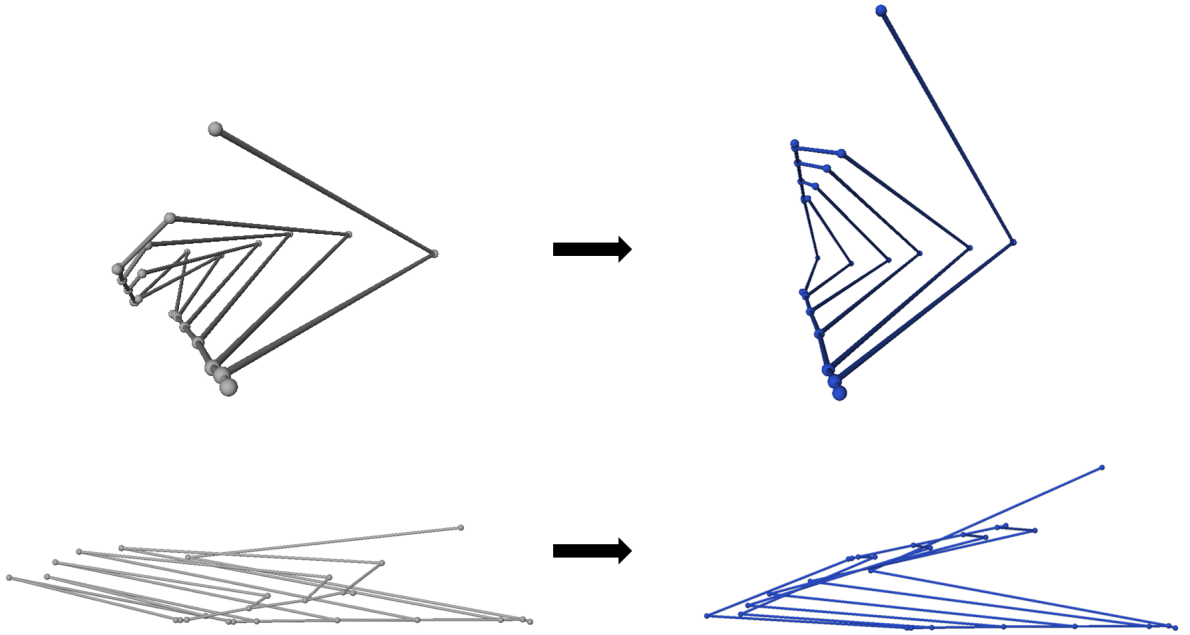


Figure 12: Virtually opening the mouth of the shark, by placing the spine nodes higher (left: original ; right: mouth opened).

Table 2: Optimization Errors for Opening Mouth in Figure 12

Point	Error in Optimization
C_h	1e-6
C_1	2e-6
C_2	2e-6
C_3	2e-6
C_4	2e-6
C_5	21.8121
P_1	1e-6
P_2	1e-6
P_3	1e-6
P_4	2e-6

Iterations: 192; Computation time: 0.017 seconds

The lower left and upper right coordinates of the 3D bounding box of the original half skeleton in Figure 12 are $(-21.39, -151.70, -994.86)$ and $(169.16, -62.71, -528.15)$, respectively.

Table 3: Rotating gill arches, hyoid arch and pharyngobranchials

Point moved	Vector “ \mathbf{v} ”	Axis “ \mathbf{e} ”
C_h	$C_h - M_h$	$A_h - B_h$
C_1	$C_1 - M_1$	$P_1 - B_1$
C_2	$C_2 - M_2$	$P_2 - B_2$
C_3	$C_3 - M_3$	$P_3 - B_3$
C_4	$C_4 - M_4$	$P_4 - B_4$
C_5	$C_5 - M_5$	$A_5 - B_5$
P_1	$P_1 - A_1$	$A_1 - A_5$
P_2	$P_2 - A_2$	$A_1 - A_5$
P_3	$P_3 - A_3$	$A_1 - A_5$
P_4	$P_4 - A_4$	$A_1 - A_5$

where M_h , M_i and M_5 are the mid-points of the lines joining A_h and B_h , A_i and $B_i \forall i \in \{1, 2, 3, 4\}$ and A_5 and B_5 , respectively.

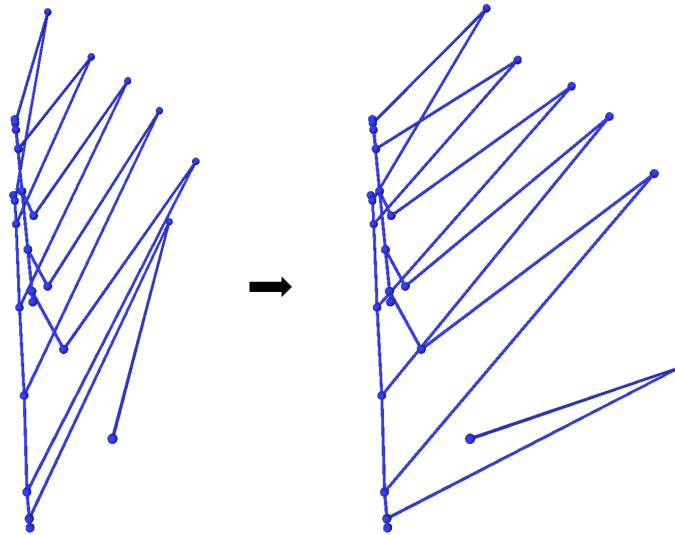


Figure 13: Rotating the gills and hyoid in order to “flare” them.

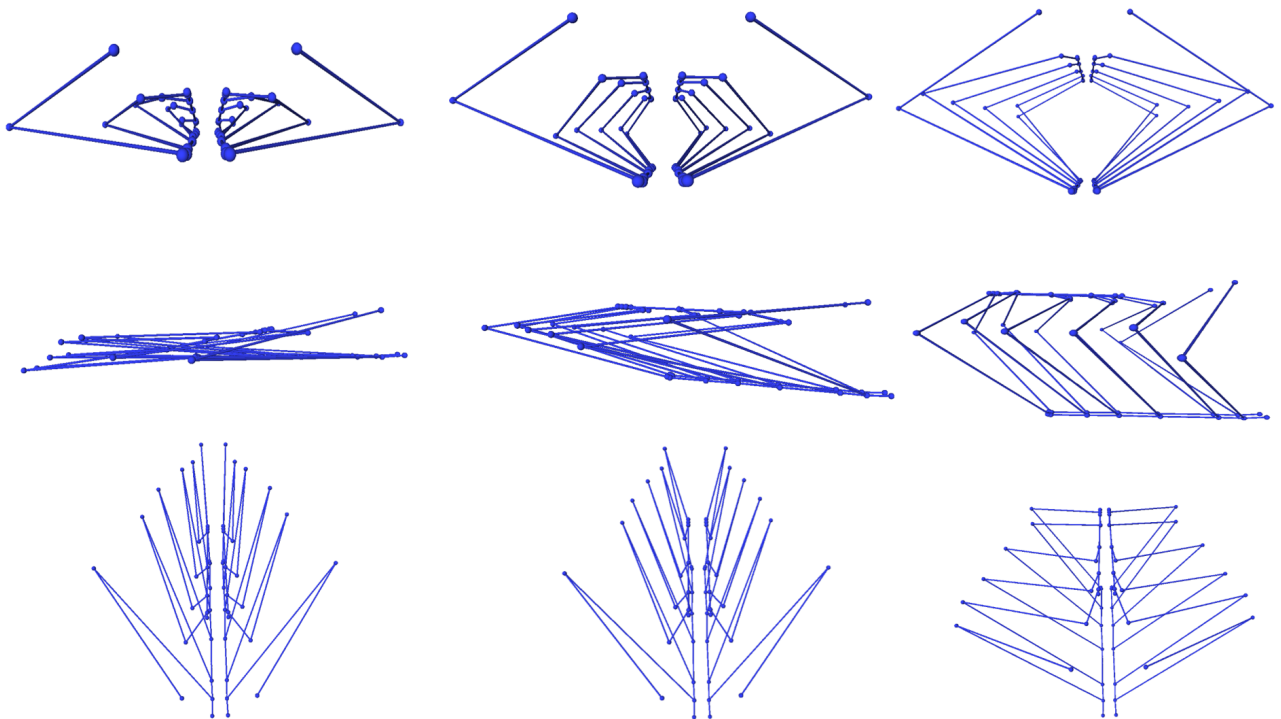


Figure 14: Examples of final skeletons in (left) closed mouth, (middle) half-open mouth and (right) open mouth configurations. The top, middle and bottom images show the front, side and top views of the skeletons, respectively. These are symmetric, in contrast to the damaged original skeletons.

3 Active Shape Model Approach

The first approach used to estimate the 3D positions of the 2D annotated landmarks is inspired by the work of Cootes and Taylor [CTCG95]. The idea is to apply the “Active Shape Model” for the 2D-to-3D shape fitting problem, which requires to solve an optimization problem to find the ideal camera parameters. Two sub-approaches, a convex and a non-convex one, have been used for this purpose. A detailed description of the original Active Shape Model and the sub-approaches will be described in this section.

3.1 The Original Active Shape Model

3.1.1 Introduction

The Active Shape Model (ASM) was first introduced by Cootes and Taylor in the early 1990s [CTCG95] and has thereafter been used intensively for the purpose of recognizing and locating non-rigid objects in the presence of noise and occlusion. It aims at providing a robust approach, which accommodates shape variability by arguing that when the objects deform in ways characteristic to the class of objects they represent, the method should be able to recognize them. Furthermore, the method is specific to a class of objects and only allows inter-shape variations, in contrast to non-rigid object recognition methods which are flexible but lack specificity [HWR⁺91, YHC92, LYO⁺90, MKW91]. This is practical in medical applications, for example, anatomical structures can vary greatly between individuals. For this, a training set is used which contains large variations of a shape and the model is allowed to deform only in ways represented in the training set by interpolating between the training shapes.

Alternative ways to model non-rigid objects include the following :

1. “Hand Crafted” models, which use simple shapes like circles and lines to model an object [YHC92, HT92].
2. Articulated Models, which use rigid objects connected by joints [BW91].
3. Active Contour Models or “Snakes”, which use energy minimizing spline curves [KWT88, HWR⁺91].
4. Finite Element methods to model objects with internal properties like elasticity [NA93].

3.1.2 Method

We follow the method from Cootes et al. [CTCG95]. The shape of a class of objects is described by a set of labeled “landmark” points, each representing a particular part of the object or its boundary. These landmark points are placed such that they represent the features necessary for identifying the object and differentiating between instances of an object. The vertices of a triangle are one example. With this in mind, a set of training shapes is created such that it includes intra-class shape variability. This is an important step as the method

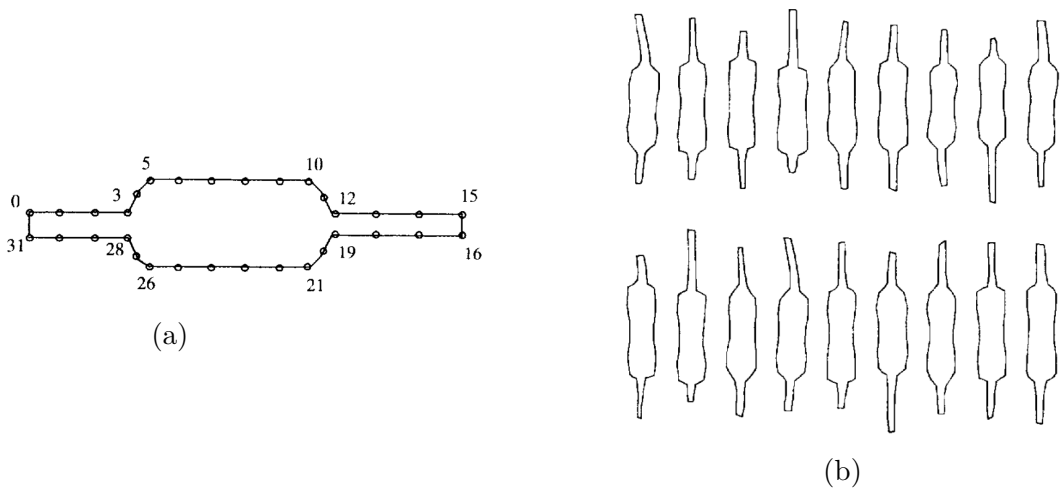


Figure 15: (a) shows a resistor shape with 32 labelled landmarks and (b) is the training set of resistor shapes [CTCG95].

only allows for variability in shape described in the training set. In a training set of hand shapes, for example, if the shape of a closed fist is not included, the method will not be able to recognize it as an instance of a hand. For an easier understanding, we will explain the method using resistor shapes. Figure 15 shows a resistor shape represented by 32 labelled landmark points and examples from the training set of resistor shapes.

As the shapes are represented by landmark points, comparison between shapes is done by comparing the corresponding points. This is achieved by the **Procrustes method** [Gow75]. Let $x_i = (x_{i1}, y_{i1}, x_{i2}, y_{i2}, \dots, x_{in}, y_{in})^T \in \mathbb{R}^{2n}$ denote the vector of the i^{th} training shape. Let $M(s, \theta)[x]$ be a rotation by θ and scaling by s , such that

$$M(s, \theta) \begin{bmatrix} x_{jk} \\ y_{jk} \end{bmatrix} = \begin{pmatrix} s \cos\theta x_{jk} - s \sin\theta y_{jk} \\ s \sin\theta x_{jk} + s \cos\theta y_{jk} \end{pmatrix} \quad (3)$$

Given two shapes x_i and x_j , aligning them amounts to finding suitable θ_j, s_j and translation $t_j = (t_{xj}, t_{yj})$, such that the squared **Procrustes distance**, d_P^2 defined below, is minimized:

$$d_P^2(x_i, x_j) = (x_i - (M(s_j, \theta_j)[x_j] + t_j))^T (x_i - (M(s_j, \theta_j)[x_j] + t_j)) \quad (4)$$

An essential step towards the formulation of this model, is aligning the entire set of training shapes. This is achieved by using **Generalized Procrustes Analysis** (GPA), described in Algorithm 4.

Note that given a set of N aligned shapes, the mean shape is computed as

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i. \quad (5)$$

Algorithm 4 Generalized Procrustes analysis for aligning a set of shapes

Require: A set of shapes, $S = \{S_1, S_2, \dots, S_k\}$, $S_i \in \mathbb{R}^{2n}$ and a threshold= tol

$ref \leftarrow S_1$

repeat

Align each S_i with ref

Compute the mean shape, $mean$, of the set of aligned shapes

$ref \leftarrow mean$

until

$d_p(ref, mean) < tol$

Upon aligning the shapes in the training set, each of them represents a point in a $2n$ -dimensional space. A cloud of N points is obtained with N training shapes and the region in which these points lie is called the “Allowable Shape Domain” [CTCG95]. New shapes can be generated by considering the points in this region as they will be broadly similar to the training shapes. Figure 16 visualizes the point cloud of resistor shapes.

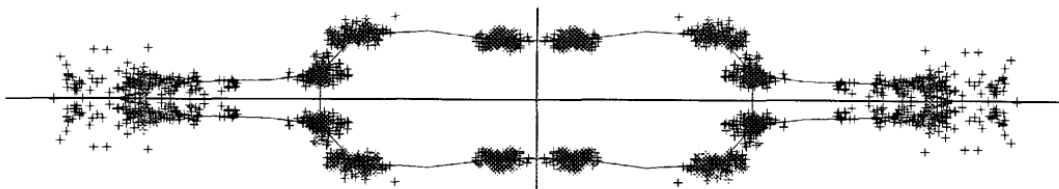


Figure 16: The outlined shape represents the mean resistor shape and crosses (+) depict the point cloud from the aligned training set. This image is taken from [CTCG95].

It is assumed that this region is a $2n$ -dimensional ellipsoid and the goal now is to compute the center, and the major axes of this ellipsoid [CTCG95]. \bar{x} corresponds to the center the latter is computed via **Principal component analysis** (PCA) on the data. Consider S to be the covariance matrix described as

$$S = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T. \quad (6)$$

The eigenvectors corresponding to the largest eigenvalues of matrix S determine

the axes of the $2n$ -dimensional ellipsoid describing most of the shape variation. It is safe to assume that the t largest eigenvectors cause most of the variation and hence the original $2n$ -dimensional ellipsoid can be approximated with a t -dimensional one.

Finally, all one has to do to access any point in the allowable shape domain is to take the mean and add a linear combination of the first t eigenvectors of the covariance matrix S

$$x = \bar{x} + Pb \quad (7)$$

where P is the matrix of the first t eigenvectors and $b = (b_1, b_2, \dots, b_t)$ is a vector of weights.

In order to visualize this, one can study how the shape of the resistors changes, with change in parameters (weights of the linear combination) in (7), refer to Figure 17.

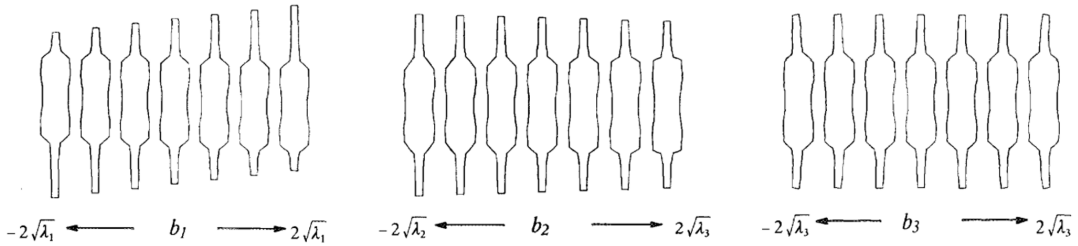


Figure 17: Effects of varying b_1, b_2 and b_3 of the resistor shape model [CTCG95].

3.2 Estimating 3D shape from 2D landmarks using ASM

3.2.1 Introduction

The first approach that we have used in order to obtain a 3D estimation of the basking shark head from 2D annotated landmarks is inspired by the active shape model described in the previous section and was proposed by Zhou et al. [ZLHD15]. Naturally, the ASM can be extended for shapes in three dimensions and the same procedure can be used as in the 2-dimensional case, described above. Hence, given a set of labelled or annotated landmarks in 2D, we can use the ASM to estimate a 3D shape whose 2D projection corresponds to the landmarks, by finding the camera parameters. In other words, the ASM can be used in a 2D-to-3D fitting problem which estimates the pose and the viewpoint parameters via an optimization problem.

3.2.2 Problem Formulation

The problem at hand is that of 2D-to-3D shape fitting, where the unknown 3D shape of an object is estimated using annotated landmarks in 2D. The landmarks are usually marked manually by the user.

Determining the 3D shape of an object from a single viewpoint is an ill-posed problem and is an impossible task without any prior information on its shape. What enables human beings from performing this task is the visual memory of the 3D shape, which can be put to use when viewing the shape from a single viewpoint. For example, one can estimate the depth of a table when viewing it from the front. In order to deal with this issue, the algorithm can be provided with a set of training shapes, as used in ASM. The unknown 3D shape can then be estimated as a linear combination of the training shapes.

Let S_1, S_2, \dots, S_k such that each $S_i \in \mathbb{R}^{3 \times p}$ be the set of training shapes where the object is described by p landmarks. Then the unknown 3D shape $X \in \mathbb{R}^{3 \times p}$ can be represented as

$$X = R \sum_{i=1}^k c_i S_i + T \mathbf{1}^T \quad (8)$$

where $R \in SO(3)$ is a rotation matrix, $T \in \mathbb{R}^{3 \times 1}$ is the translation vector and $\mathbf{c} = (c_1, \dots, c_k)$ is the vector of weights of the linear combination.

Let $W \in \mathbb{R}^{2 \times p}$ represent the annotated 2D landmarks, then

$$W = \prod (R \sum_{i=1}^k c_i S_i + T \mathbf{1}^T), \quad (9)$$

where \prod corresponds to the weak-perspective camera matrix [Alo90] such that

$$\prod = \begin{pmatrix} s & 0 & 0 \\ 0 & s & 0 \end{pmatrix}$$

with $s \in \mathbb{R}$. This is a widely used assumption in order to reduce the problem complexity and works well in most cases where the object depth is small compared to the distance of the object from the camera.

The data is centered to simplify the problem:

$$\begin{aligned} W &= \begin{pmatrix} s & 0 & 0 \\ 0 & s & 0 \end{pmatrix} R \sum_{i=1}^k c_i S_i \\ &= s R_{[1,2]} \sum_{i=1}^k c_i S_i. \end{aligned}$$

Absorbing s in \mathbf{c} , i.e. assuming $\mathbf{c} = (sc_1, sc_2, \dots, sc_k)$, gives

$$W = R_{[1,2]} \sum_{i=1}^k c_i S_i \quad (10)$$

In order to solve for the pose parameters (\mathbf{c}) and the viewpoint parameter (R), the following non-convex optimization problem needs to be solved, which minimizes the re-projection error and induces sparsity in the vector of weights:

$$\begin{aligned} \min_{R_{[1,2]}, \mathbf{c}} \quad & \|W - R_{[1,2]} \sum_{i=1}^k c_i S_i\|_F^2 + \lambda \|\mathbf{c}\|_1 \\ \text{s.t.} \quad & R_{[1,2]} R_{[1,2]}^T = I_2, \end{aligned} \quad (11)$$

where $\|\cdot\|_F^2$ refers to the squared Frobenius norm of a matrix and $\|\cdot\|_1$ to be l_1 -norm of a vector. The second term can be omitted if sparsity of \mathbf{c} is not required. We solve (11) using two different approaches.

3.2.3 Non-convex Formulation

The first approach aims to solve (11), a non-convex optimization problem due to the orthogonality constraint. For this, an alternating minimization scheme can be employed, which updates $R_{[1,2]}$ and \mathbf{c} alternatively via optimization over Stiefel manifold, $V_{2,3} = \{Q \in \mathbb{R}^{2 \times 3}, QQ^T = I_2\}$, and \mathbb{R}^k respectively. In this case, however, the optimization is sensitive to initialization.

Zhou et al. describe a method to convert the optimization problem in (11) to a convex problem, by relaxing the orthogonality constraint. The next section provides the details of this formulation, adopted from the publication by Zhou et al. [ZLHD15].

3.2.4 Convex formulation

This section deals with the convex relaxation of the orthogonality constraint in (11) and converting the problem to a convex optimization problem.

Note that in (10), a single rotation is applied to a linear combination of the training shapes. If, however, a separate rotation is applied to each training shape, a linear representation of W can be reached and the bi-linear form in (10) is removed. As the degrees of freedom are increased, this formulation may lead to bizarre shapes which are very different from the ones in the training set. This issue can be dealt with to a large extent by including a sparsity constraint on the number of training shapes used.

The new formulation of (11) is:

$$\begin{aligned}
W &= \prod_{i=1}^k (\sum_{i=1}^k c_i (R_i S_i)) \\
&= \sum_{i=1}^k c_i R_{i[1,2]} S_i \\
&= \sum_{i=1}^k M_i S_i \\
&\text{s.t. } M_i M_i^T = c_i I_2
\end{aligned} \tag{12}$$

where $M_i = c_i R_{i[1,2]} \in \mathbb{R}^{2 \times 3}$.

The next step towards achieving a convex formulation of (12) is to replace the orthogonality constraint by its convex counterpart. The convex hull of a set X , denoted by $\text{conv}(X)$, is the smallest convex set containing X . Zhou et al. [ZLHD15] prove that

$$\text{conv}\{Y \in \mathbb{R}^{m \times n} \mid Y^T Y = s^2 I_n\} = \{Y \in \mathbb{R}^{m \times n} \mid \|Y\|_2 \leq |s|\},$$

where $\|M_i\|_2$ refers to the spectral norm of M_i , which is its largest singular value. Using this result, we are only a few steps away from achieving the final formulation of the optimization problem introduced in (11), which uses sparsity of weights and the relaxed orthogonality constraint.

Consider the following optimization problem

$$\begin{aligned}
&\min_{\mathbf{c}, M_1, \dots, M_k} \sum_{i=1}^k |c_i| \\
&\text{s.t. } W = \sum_{i=1}^k M_i S_i, \\
&\|M_i\|_2 \leq |c_i| \forall i \in [1, k]
\end{aligned} \tag{13}$$

We can rewrite (13) as

$$\begin{aligned}
&\min_{M_1, \dots, M_k} \sum_{i=1}^k \|M_i\|_2 \\
&\text{s.t. } W = \sum_{i=1}^k M_i S_i.
\end{aligned}$$

The above formulation works in noiseless cases, but to account for noise in real-world applications, the following regularized least-squares optimization problem can be used

$$\min_{M_1, \dots, M_k} \|W - \sum_{i=1}^k M_i S_i\|_F^2 + \lambda \sum_{i=1}^k \|M_i\|_2. \quad (14)$$

The solution of (14) will estimate M_i minimizing the objective. Note that minimizing the spectral norm of a matrix is equivalent to minimizing the l_∞ -norm of the vector of its singular values [ZLHD15]. Doing so reduces the matrix norm towards zero, which in turn shrinks its singular values to be equal. This forces the matrix towards a zero matrix, which aims at inducing sparsity of the weights and orthogonality of each M_i . Each M_i can then be used to estimate each c_i and $R_{i[1,2]}$ using

$$\begin{aligned} c_i &= \|M_i\|_2, \\ R_{i[1,2]} &= \frac{M_i}{c_i} \end{aligned}$$

The third row the the rotation matrix R_i is calculated by taking the cross-product of the first two rows. The final shape is estimated by:

$$X = \sum_{i=1}^k c_i R_i S_i$$

Solving the optimization problem

The algorithm to solve the optimization problem in (14) is based on the Alternating Direction Method of Multipliers (ADMM) [BPC11]. By introducing an auxiliary variable Y , (14) can be reformulated as:

$$\begin{aligned} \min_{M_c, Y} \frac{1}{2} \|W - Y S_c\|_F^2 + \lambda \sum_{i=1}^k \|M_i\|_2, \\ \text{s.t. } M_c = Y, \end{aligned} \quad (15)$$

where $M_c \in \mathbb{R}^{2 \times 3k}$ such that $M_c = [M_1 M_2 \dots M_k]$, and $S_c \in \mathbb{R}^{3k \times p}$ such that $S_c = [S_1 S_2 \dots S_k]^T$ (the subscript c refers to concatenated).

The augmented Lagrangian of (15) is

$$\begin{aligned} L(M_c, Y, D) &= \frac{1}{2} \|W - Y S_c\|_F^2 + \lambda \sum_{i=1}^k \|M_i\|_2 \\ &+ \langle D, M_c - Y \rangle + \frac{\mu}{2} \|M_c - Y\|_F^2, \end{aligned} \quad (16)$$

where D is the dual variable and μ is the step size parameter. The updates of each variable at time-step t according to ADMM are

$$M_c^t = \arg \min_{M_c} L(M_c, Y^{t-1}, D^{t-1}) \quad (17)$$

$$Y^t = \arg \min_Y L(M_c^t, Y, D^{t-1}) \quad (18)$$

$$D^t = D^{t-1} + \mu(M_c^t - Y^t)$$

until convergence is reached. Expanding (17), we get

$$\begin{aligned}
\arg \min_{M_c} L(M_c, Y^{t-1}, D^{t-1}) &= \arg \min_{M_c} \left(\frac{\lambda}{\mu} \sum_{i=1}^k \|M_i\|_2 + \frac{1}{\mu} \langle M_c - Y^{t-1}, D^{t-1} \rangle \right. \\
&\quad \left. + \frac{1}{2} \|M_c - Y^{t-1}\|_F^2 \right) \\
&= \arg \min_{M_c} \frac{1}{2} \left(\|M_c - Y^{t-1}\|_F^2 + 2 \langle M_c - Y^{t-1}, \frac{D^{t-1}}{\mu} \rangle \right. \\
&\quad \left. + \left\| \frac{D^{t-1}}{\mu} \right\|_F^2 \right) + \frac{\lambda}{\mu} \sum_{i=1}^k \|M_i\|_2 \\
&= \arg \min_{M_c} \frac{1}{2} \left\| M_c - Y^{t-1} + \frac{1}{\mu} D^{t-1} \right\|_F^2 + \frac{\lambda}{\mu} \sum_{i=1}^k \|M_i\|_2 \\
&= \arg \min_{M_1, M_2, \dots, M_k} \sum_{i=1}^k \left(\frac{1}{2} \|M_i - Q_i^{t-1}\|_F^2 + \frac{\lambda}{\mu} \|M_i\|_2 \right)
\end{aligned} \tag{19}$$

where Q_i^{t-1} is the i^{th} column matrix of $Y^{t-1} - \frac{1}{\mu} D^{t-1}$. The proximal problem [ZLHD15] can be used to solve (19) to get

$$M_i^t = \mathcal{D}_{\frac{\lambda}{\mu}}(Q_i^{t-1}) \quad \forall i \in [1, k].$$

where $\mathcal{D}_{\frac{\lambda}{\mu}}(Q_i^{t-1})$ is the solution to the proximal problem.

(18) being a quadratic form of Y has the following solution

$$Y^t = (W S_c^T + \mu M_c^t + D^{t-1})(S_c S_c^T + \mu I)^{-1}.$$

This concludes the algorithm.

3.3 Implementation

Section 2 describes the creation of piece-wise linear skeletons of basking sharks, using real-world data in form of CT scans. These skeletons are an abstract representation of the anatomical skeleton, and are 3D spatial graphs, consisting of 52 nodes and 58 segments. The latter describes the connection between the nodes, and is the same for every skeleton. Hence, each skeleton is uniquely identified by its set of nodes. A total of 27 skeletons were created in various configurations and we consider these to be our set of training shapes, $\{S_i\}_{i=1}^{27}$, $S_i \in \mathbb{R}^{3 \times 52} \forall i$ ($p = 52$ and $k = 27$). Given a set of 2D annotated landmarks, $W \in \mathbb{R}^{2 \times 52}$ (centered), we estimate the 3D shape corresponding to these using the two sub-approaches in Sections 3.2.3 and 3.2.4.

For the optimization in (11), we omit the second term since sparsity of the vector of weights is not of interest to us. The reason for this is that the opening of

the shark’s mouth is a movement which interpolates between the configurations created (closed, half open and open mouth), hence all shapes can contribute to the unknown shape. We then find the rotation R and the weights \mathbf{c} via alternatively minimizing each parameter. The rotation is updated via steepest descent [C+47] on Stiefel manifold, $V_{2,3}$, and the weights are updated via steepest descent on \mathbb{R}^{27} . This was performed in Python using the Pymanopt [BMAS14] library, which includes optimization routines on manifolds. The gradients were computed automatically using the Autograd [MDA15] library which uses reverse-mode differentiation to compute gradients.

The second sub-approach aims to solve (14) via convex optimization and was implemented in Amira, using C++. The values of λ and μ in (15) were considered to be 0.05 and 0.5 respectively, and the variables were updated until convergence (tolerance=1e-05).

The results obtained are visualized and compared in Section 5.

4 Kendall’s Shape Space Approach

The second method used to estimate the 3D position of 2D annotated landmarks, uses the theory of Kendall’s shape space. It has been used widely in the context of geometric morphometrics, a field whose significance is ever increasing due to its applications in a variety of fields, including bio-medical sciences, anthropology and image processing. The tricky part when it comes to studying shape variability in shape spaces like Kendall’s Shape Space, is the high dimensionality and non-linearity of these spaces, as they usually lie in a highly-curved region. A useful approach is to consider geodesics on this high-dimensional, curved shape space. Note that by “shape-space”, we mean an abstract representation of a space, where each point defines a unique shape. The distance between the points in this space refers to a measure of difference between the shapes represented by these points. In this section, we provide an introduction to Kendall’s Shape Space, and define the necessary computations on it. Then, we describe how it can be applied to the 2D-to-3D shape fitting problem. One may refer to [HH14, Pen06], for more information on statistical analysis on Riemannian manifolds.

4.1 Kendall’s Shape Space

An intuitive interpretation of a “shape” was provided by the English statistician and mathematician David G. Kendall in the late 1900s [Ken84]. According to this, a “shape” in \mathbb{R}^n , is a set of p points in \mathbb{R}^n , with translations, rotations and scale removed. Equivalently, a shape is the relative arrangement of a given number of landmarks in \mathbb{R}^n . With this definition, statistical estimates like average and variations in shape can be computed in the shape space.

Let $x_1, x_2, \dots, x_p \in \mathbb{R}^n$ and construct $x = [x_1 \ x_2 \ \dots \ x_p] \in \mathbb{R}^{n \times p}$ by horizontally stacking the x_i s. Let \bar{x} denote the Euclidean mean of these points and define the “centroid size” [Boo97] or simply “size” of x by

$$size(x) = \sqrt{\sum_{i=1}^p (x_i - \bar{x})^2}. \quad (20)$$

Replacing x by $\frac{x - \bar{x}}{size(x)}$, i.e., subtracting the mean from each x_i and dividing by the size, implies that x lies on the unit sphere \mathcal{S}_p^n , where

$$\mathcal{S}_p^n = \{x \in \mathbb{R}^{n \times p} : \sum_{i=1}^p x_i = 0, \|x\| = 1\}. \quad (21)$$

The set \mathcal{S}_p^n is called the “pre-shape space” and contains every point in $\mathbb{R}^{n \times p}$ with translations and scale removed (i.e., with $size=1$). The term pre-shape space is used instead of just shape space as the rotations are not yet removed. Upon removing the rotations from the pre-shape space, we obtain our shape

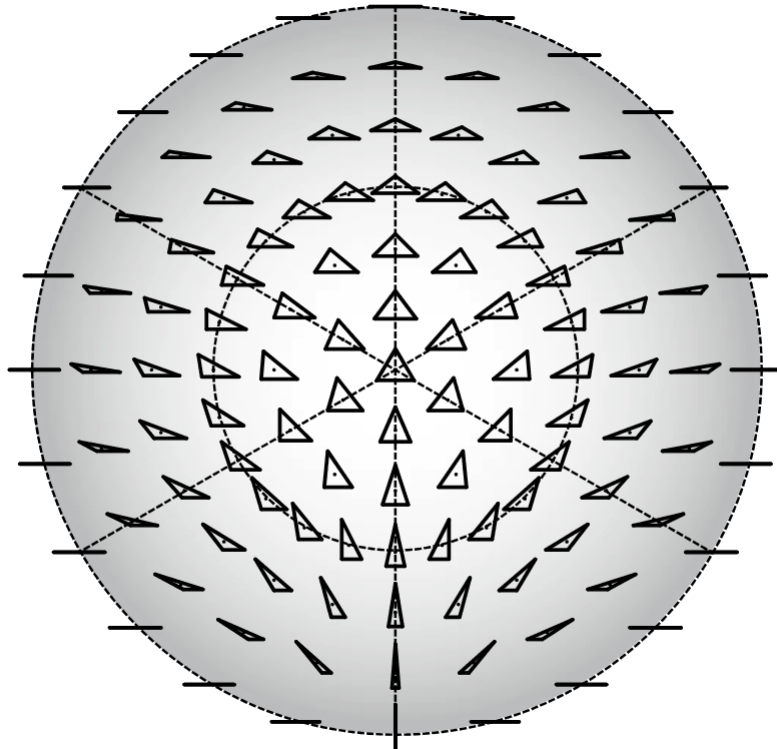


Figure 18: Being curved, multi-dimensional spaces, Kendall shape spaces are hard to visualize. The shape space of triangles in 2D, i.e. Σ_3^2 , is relatively easier to visualize as it is the surface of a sphere in 3- dimensions. This figure shows the view from the north pole of the sphere, which is visualized as the equilateral triangle. The meridians and the equator correspond to the isosceles and flat triangles, respectively [Kli20]

space. This is called Kendall's shape space, and is represented by the quotient space $\Sigma_p^n = \mathcal{S}_p^n / \sim$ where $x \sim y \iff \exists R \in SO(n)$ with $x = Ry$. This is a Riemannian manifold of equivalence classes. Provided that $p \geq n + 1$, the dimension of this space can be computed by subtracting the dimensions lost from removing scaling (1 dimension), translations (n dimensions) and rotation ($(n-1)/2$ dimensions). Thus, the dimension of the space Σ_p^n is $n(p-1) - \frac{1}{2}n(n-1) - 1$.

Let π denote the canonical projection of \sim , then the distance between two shapes, $\pi(x)$ and $\pi(y)$ is

$$d_\Sigma(x, y) = \arccos\left(\sum_{i=1}^n \lambda_i\right). \quad (22)$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{m-1} \geq |\lambda_m|$ are the pseudo-singular values of yx^T [NYHSvT20].

For every $x, y \in \mathcal{S}_p^n$, there exists an optimal rotation, $R \in SO(n)$, such that:

$$d_\Sigma(x, y) = d(x, Ry) \quad (23)$$

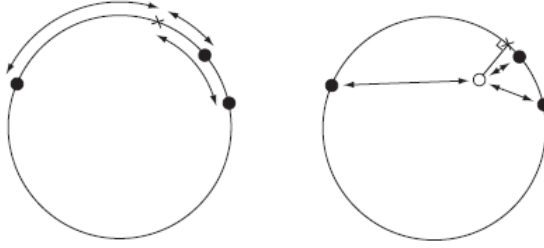


Figure 19: The intrinsic mean (left) and the extrinsic mean (right) of three points on S^1 , represented by crosses [BWHK07].

The points x and y are said to be *well-positioned* (denoted by $x \stackrel{w}{\sim} y$), if and only if, yx^T is symmetric and $d(x, y) = d_\Sigma(x, y)$.

4.1.1 The Fréchet Mean

An important measure when studying the statistics of non-linear spaces is the “intrinsic mean” or Fréchet mean. It is the generalization of the Euclidean mean to non-linear spaces, and a useful tool for calculating means on manifold valued data. Let M be a Riemannian manifold M , defined using the metric d_M , defined as the geodesic distance between two points on M . The Fréchet mean of a set of k points, $\{x_1, x_2, \dots, x_k\}$ on M is defined as the point on M which minimizes the sum of squared distances to this set of points, that is:

$$\bar{x}_F = \underset{y \in M}{\operatorname{argmin}} \sum_{i=1}^k d_M(x_i, y)^2. \quad (24)$$

Note that for a manifold M embedded in a Euclidean space, it is possible to calculate the so-called “extrinsic mean”, which is defined as the projection of the Euclidean mean on the manifold and is equivalent to solving the following minimization problem [BWHK07]:

$$\bar{x}_E = \underset{y \in M}{\operatorname{argmin}} \sum_{i=1}^k |x_i - y|^2. \quad (25)$$

In case of Kendall’s Shape Space, we are interested in calculating the Fréchet mean in order to obtain an average shape of a set of shapes. For the points $\pi(x_1), \pi(x_2), \dots, \pi(x_k) \in \Sigma_p^n$, it is defined as:

$$\mu_F(\pi(x_1), \pi(x_2), \dots, \pi(x_k)) = \underset{\pi(y) \in \Sigma_p^n}{\operatorname{argmin}} \left(\sum_i d_\Sigma(y, x_i)^2 \right) \quad (26)$$

This definition can easily be extended to that of a weighted Fréchet mean, analogous to a weighted mean in Euclidean terms, as:

$$\mu_F(\mathbf{c}; \pi(x_1), \pi(x_2), \dots, \pi(x_k)) = \underset{\pi(y) \in \Sigma_p^n}{\operatorname{argmin}} \left(\sum_i c_i d_\Sigma(y, x_i)^2 \right) \quad (27)$$

where $\mathbf{c} = (c_1, c_2, \dots, c_k)$ is a vector of weights.

The next section explains the details of applying Kendall’s shape space to the 2D-to-3D fitting problem.

4.2 Problem Formulation

We will now formulate a 2D-to-3D fitting optimization problem which, given a set of p annotated landmarks in 2D, aims to find the shape in Kendall’s shape space Σ_p^n , whose projection under the weak perspective camera model is closest to this set of 2D landmarks. For this, we use a set of training shapes in Σ_p^n , and try to estimate the unknown 3D shape as a weighted Fréchet mean of these shapes. This differs from the first approach, where the unknown shape is estimated to simply be a linear combination of training shapes in 3D.

Let $W = [w_1, \dots, w_p] \in \mathbb{R}^{2 \times p}$ represent the annotated 2D landmarks, S_1, \dots, S_k , such that each $S_i \in \Sigma_p^n$ be the set of training shapes and $\mathbf{c} = (c_1, \dots, c_k)$ be the vector of weights. “Normalise” W such that it is centered and $size(W) = 1$. Then the 3D shape corresponding to W , $X \in \mathbb{R}^{3 \times p}$, is estimated by solving the following constrained least-squares optimization problem:

$$\begin{aligned} \min_{\mathbf{c}, R} \quad & \frac{1}{2} \|W - R\mu_F(\mathbf{c}; S_1, S_2, \dots, S_k)\|_F^2 \\ \text{s.t.} \quad & RR^T = I_2, \end{aligned} \tag{28}$$

where $R \in V_{2,3} = \{Q \in \mathbb{R}^{2 \times 3} : QQ^T = I_2\}$ is the projection of $\mu_c(S_1, S_2, \dots, S_k)$ under the weak perspective camera model [Alo90]. The 3D shape is estimated as the weighted Fréchet mean, $\mu_F(\mathbf{c}; S_1, \dots, S_k) \in \Sigma_p^n$. It is important to note that as the 2D landmarks are normalised, the second term in the Frobenius norm in (28) must also be normalised for a meaningful comparison.

4.3 Solving the Optimization Problem

We solve the constrained least-squares optimization problem in (28) via an alternating minimization scheme, which updates the two parameters \mathbf{c}, R alternately. First, \mathbf{c} is fixed and R is updated via manifold optimization on $V_{2,3}$. Then, R is fixed and \mathbf{c} is updated via optimization on \mathbb{R}^k . These steps are alternated until convergence. This optimization is sensitive to initialisation, due to the non-convexity of the orthogonality constraint.

4.4 Implementation

We test Kendall’s Shape Space approach for estimating the 3D shape of the head skeleton of a basking shark, given 2D annotated landmarks of the same. The implementation is performed using Python. In section 2, we created 27 piecewise linear skeletons of basking sharks, each represented uniquely by a set of 52 nodes in 3D. We use these to create the training shapes which will be used in the optimization in (28).

The first step is to remove the translation (centering the points) and scale (converting *size* to 1) from the set of nodes of each skeletons. By doing so, they are converted to pre-shapes, denoted by S_1, S_2, \dots, S_{27} . As the rotations are not removed, they cannot yet be considered shapes, which are equivalence classes of pre-shapes. We can, however, perform computations on the latter, as if two points are well-positioned, the pre-shape distance between them is equivalent to the shape distance.

We consider S_1, S_2, \dots, S_{27} to be the training shapes for estimating the 3D shape of a set of annotated 2D landmarks, W , and solve the optimization problem in (28) to obtain a weighted Fréchet mean of these training shapes, such that its projection under the weak perspective model is closest to W . We do this by performing alternating minimization of the parameters R and \mathbf{c} . The former is updated via steepest descent [C⁺47] on $V_{2,3}$ and the latter via steepest descent on \mathbb{R}^{27} , using Pymanopt [BMAS14], a manifold optimization library and automatic differentiation using Autograd [MDA15]. For computing the Fréchet mean of training shapes, we use the recursive estimator presented by Chakraborty et al [CBMV18], and the library `Geomstats` [MGLB⁺20].

The results are visualized in Section 5.

5 Results

This section presents the results obtained from the ASM and KSS approaches, applied to the 2D-to-3D shape fitting problem. As explained in Sections 3 and 4, both approaches use a set of pre-defined 3D training shapes to estimate the 3D shape corresponding to a set of 2D annotated landmarks. The former estimates it simply as a linear combination of the training shapes, and the latter as a weighted Fréchet mean of the training shapes on Kendall’s shape space. For this, the optimization problems in (11), (14) and (28) are solved to obtain the weights and projection parameters. We consider the shape of the parts of the head skeleton of a basking shark which are involved in its feeding motion. The shape consists of 52 ordered landmarks in 3D, and we use a set of 27 training shapes. Section 2 gives the details on creating them from real-world data in form of CT scans of basking sharks.

To judge the exactness of recovery, we perform a “cross-validation” or “leave one out” study, which excludes a shape from the set of training shapes and tries to reconstruct it using the other training shapes. Then, to study robustness, we consider the performance of the approaches in presence of noise in the 2D landmarks. Finally, performance on real-world images of basking sharks is compared. The average computation time based on derived prototype implementation (see Sections 3.3 and 4.4) is:

- The ASM non-convex approach takes ≈ 7 mins for 50 iterations.
- The ASM convex approach takes ≈ 0.65 seconds until convergence.
- The KSS approach takes ≈ 1 hour for 50 iterations.

Experiments were run on a computer with Intel(R) Xeon(R) CPU E5-1650 v3 at 3.50GHz, 64GB RAM and GeForce GTX 980 Ti GPU.

5.1 Exactness of recovery

The first test to compare the two approaches is the “leave one out” test, which excludes one shape from the set of training shapes, and tries to estimate it using the other training shapes. The shape distance is then computed as the Procrustes distance between the true shape and estimated shape, which is defined as the square root of the sum of differences between corresponding points in both shapes, when aligned using the Procrustes algorithm.

For the i^{th} training shape, $S_i \in \mathbb{R}^{3 \times 52}$, let $W_i \in \mathbb{R}^{2 \times 52}$ represent its 2D projection. This projection can be bi-laterally symmetric (i.e., there exists a line of symmetry which divides the set of projected points in two mirrored halves) or asymmetric, depending on the plane on which the landmarks are projected. Conceptually, a good approach is one which gives a bi-laterally symmetric estimated 3D shape, no matter what the symmetry of the 2D landmarks is. Let X_i be the shape estimated from the other training shapes, for the projection W_i of the shape S_i . The violin plots in Figures 20 and 22 show the distance between the shapes S_i and X_i for symmetric and asymmetric projections, respectively. Some visualizations of the skeletons can be seen in Figures 21 and 23.

Table 4: Shape distance statistics for symmetric projection

	ASM non-convex	ASM convex	KSS
Mean	0.107	0.516	0.097
Variance	0.003	0.129	0.003

Table 5: Shape distance statistics for asymmetric projection

	ASM non-convex	ASM convex	KSS
Mean	0.213	0.354	0.066
Variance	0.095	0.032	0.002

In case of symmetric projection, we make the following observations:

- The ASM non-convex and KSS approaches clearly out-perform the ASM convex approach for all training shapes. The latter produces shapes which align perfectly with the 2D landmarks, but appear very different from the training shapes.
- For comparing the ASM non-convex and KSS, a paired two-sample t -test was performed on their samples of estimation errors. A p -value of 0.30 was obtained, which is not significant.
- The estimated 3D shape appears symmetric for all the approaches and its scale is comparable to that of the training shapes.

In case of asymmetric projection, we make the following observations:

- KSS approach greatly out-performs both the ASM approaches.
- The estimated 3D shapes in the case of ASM non-convex and KSS approaches, appear bi-laterally symmetric. This is not true for the ASM convex approach, which estimates an asymmetric 3D shape in most cases. Moreover, for 10 of the training shapes ($\approx 37\%$ of the cases), the estimated 3D shape is squished and lacks depth, but still aligns well with the 2D landmarks. The scale of the estimated shape is comparable to the training shapes in all the approaches.
- Upon performing a two-sample t -test on the samples of estimation errors obtained from the ASM non-convex and KSS approaches, a p -value of 0.02 was obtained, which is significant.

5.2 Robustness

The second criteria by which we compare the approaches is robustness, which is measured by their performance when dealing with noisy data. A good approach is one which is able to estimate the true 3D shape, even when the 2D landmarks

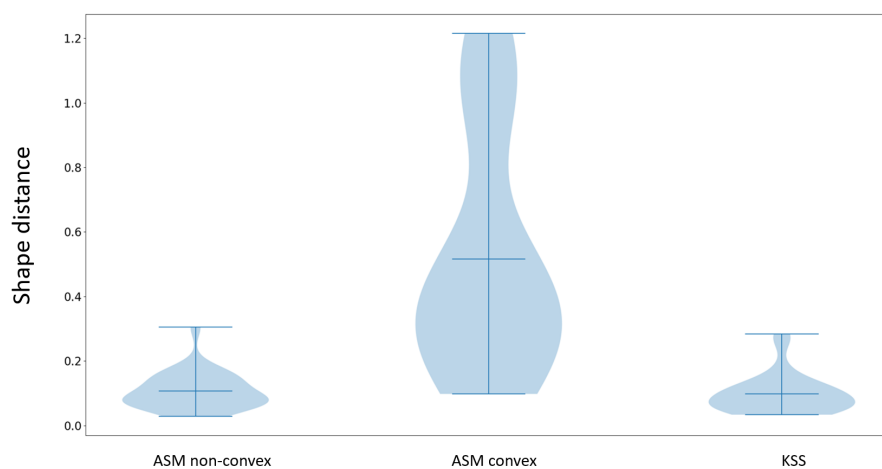


Figure 20: Shape distances between true shapes S_i and estimated shapes X_i , for a symmetric projection.

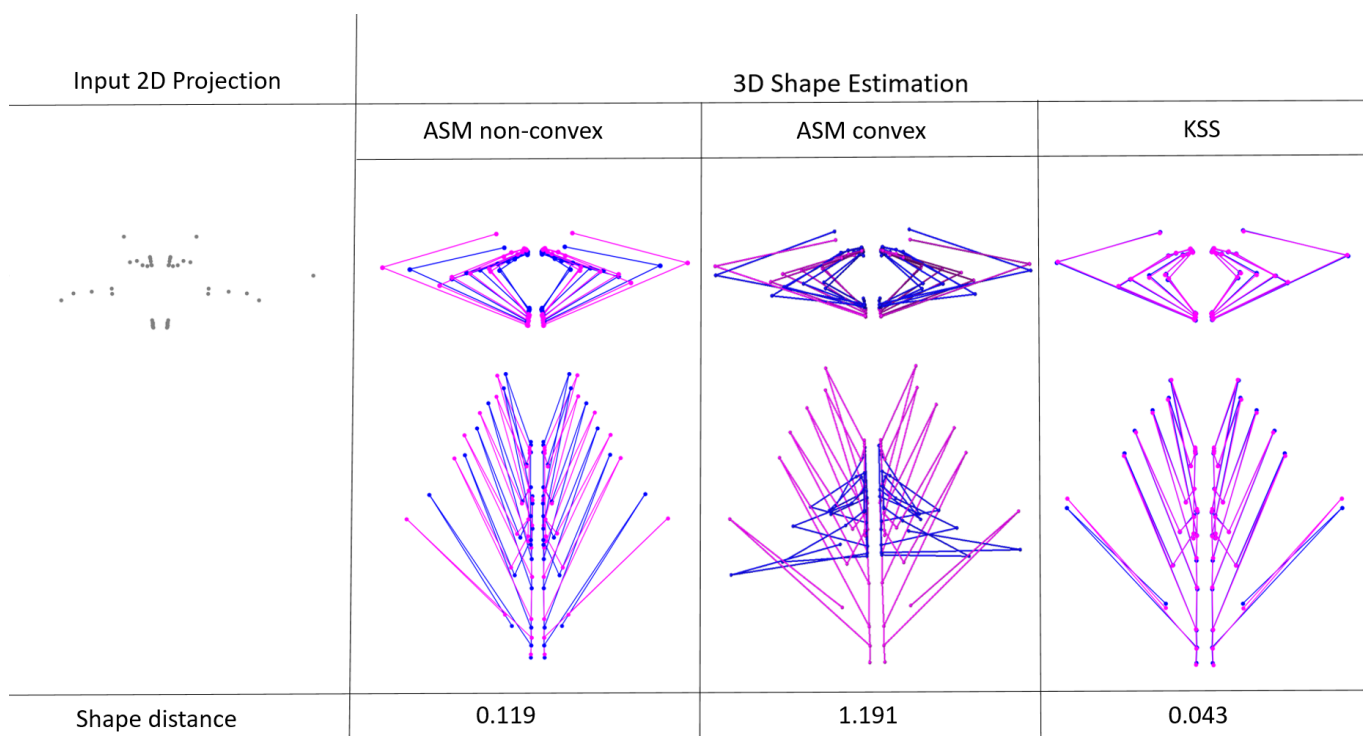


Figure 21: The front and top views of the estimated 3D shapes, in case of symmetric 2D projection. The true shape and estimated shapes are represented by the magenta and blue skeletons, respectively.

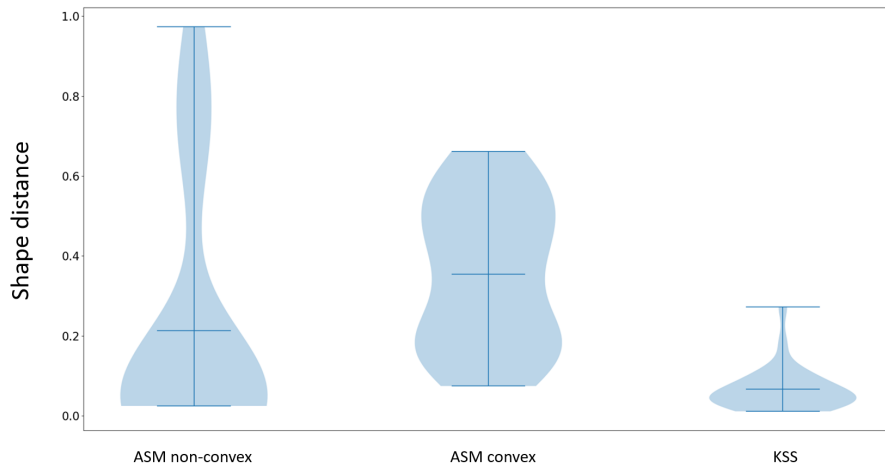


Figure 22: Shape distances between true shapes, S_i , and estimated shapes, X_i , for asymmetric projection.

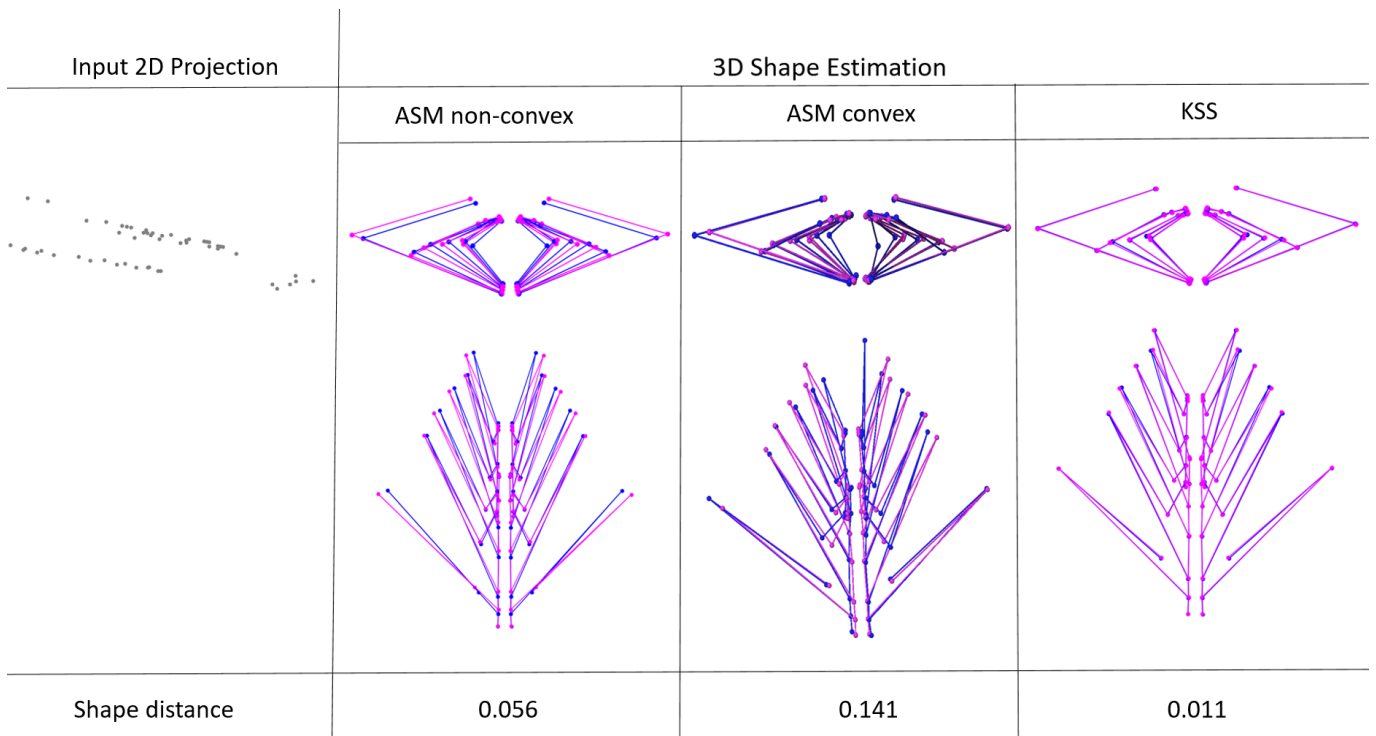


Figure 23: The front and top views of the estimated 3D shapes, in case of asymmetric 2D projection. The true and estimated shapes are represented by the magenta and blue skeletons, respectively.

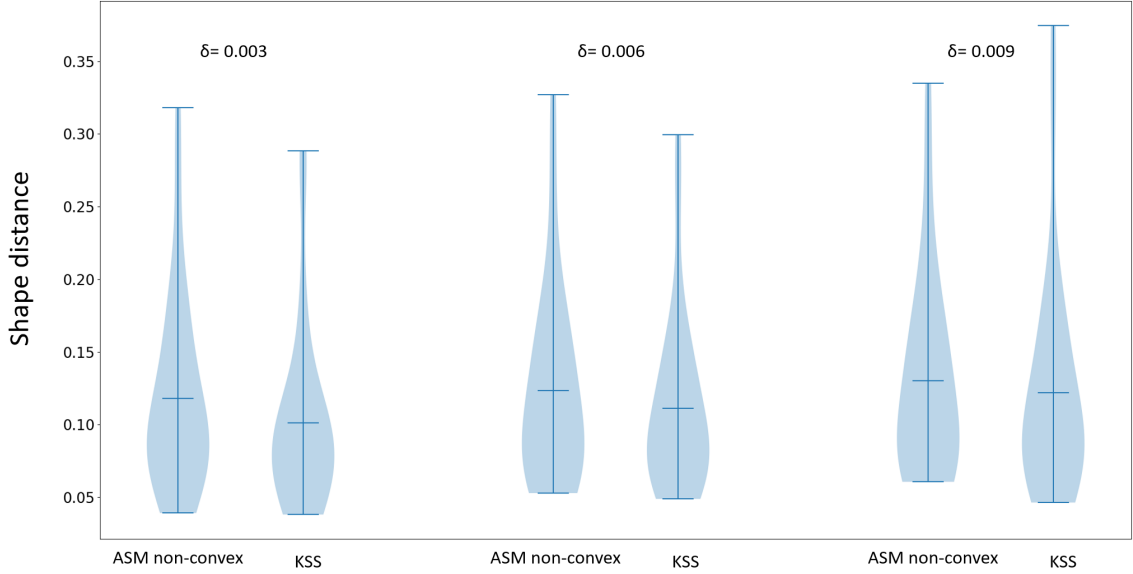


Figure 24: Shape distance between true shape S_i and estimated shape X_i , when noise sampled from $\mathcal{N}(0, s_i * \delta)$ is added to the symmetric projection.

contain some noise. This is useful in a real-world scenario, where the data usually contain some amount of noise. This could be in form of errors when placing landmarks on an image manually, especially when the resolution of the image is low and/or the positions where the landmarks are to be placed are not clearly visible. The pharyngobranchials, for example, are usually not clearly visible in most images of basking sharks. The task becomes even harder, when the placement of landmarks is performed by a non-professional. We test this on each training shape, by adding noise to its symmetric 2D projection and estimating the 3D shape using the other training shapes, i.e., leaving out the shape being estimated like in section 5.1. The shape distance between the estimated 3D shape and the true shape is then compared. The ASM non-convex and KSS approaches will be compared in this section. For the scope of this master’s thesis, we decided to leave out the ASM convex approach as its performance for the noiseless cases in Section 5.1 was not satisfactory, which makes it unreliable for the problem at hand.

Let $W_i \in \mathbb{R}^{2 \times p}$ denote the symmetric 2D projection of training shape S_i , such that $size(W_i) = s_i$ (see (20)). Let $N \in \mathbb{R}^{2 \times p}$ be a matrix of noise, such that N_{jk} is sampled from $\mathcal{N}(0, s_i * \delta)$, the normal distribution centered at 0 with scale $s_i * \delta$, $\delta \in \mathbb{R}^+$. We denote the noisy 2D landmarks by $\tilde{W}_i = W + N \in \mathbb{R}^{2 \times p}$ and the estimated 3D shape by \tilde{X}_i . The violin plots in Figure 24 compare the shape distances between S_i and \tilde{X}_i for the approaches, for different values of δ . Some examples are visualized in Figure 25.

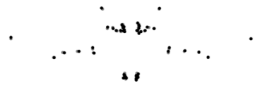
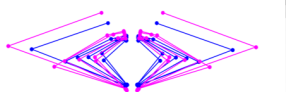
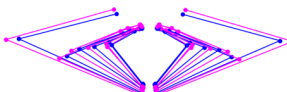
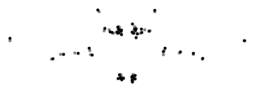
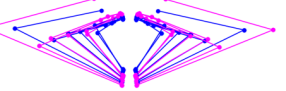
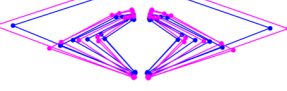
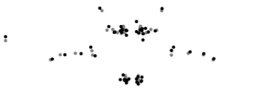
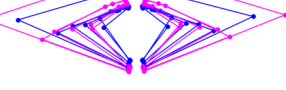
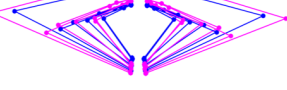
2D Projection	3D Shape Estimation	
	ASM non-convex	KSS
		
Shape distance	0.136	0.069
		
Shape distance	0.146	0.094
		
Shape distance	0.155	0.120

Figure 25: The front and top views of the estimated 3D shapes, in case of noisy 2D landmarks, for (top) $\delta = 0.003$, (middle) $\delta = 0.006$ and (bottom) $\delta = 0.009$. The gray and the **black** nodes represent the noiseless and noisy projections, respectively. The true and estimated shapes are represented by the magenta and blue skeletons, respectively.

Table 6: Shape distance statistics for noisy 2D landmarks

δ	0.003		0.006		0.009	
	ASM NC	KSS	ASM NC	KSS	ASM NC	KSS
Mean	0.117	0.100	0.123	0.111	0.130	0.121
Variance	0.004	0.003	0.004	0.003	0.004	0.005

Table 7: p -values for two-sample t -test, in case of noisy 2D landmarks.

δ	0.003	0.006	0.009
p -value	0.06	0.14	0.44

In case of noisy data, we have the following observations:

- The KSS approach out-performs the ASM non-convex approach when estimating the true shape in the presence of noisy 2D annotated landmarks. Table 6 summarizes some statistics obtained from the samples.
- Table 7 shows the p -values were obtained upon performing a two sample t -test on the samples of their shape errors. The difference between samples is not significant in any of the cases.
- The estimated shape in both cases is symmetric and its scale is comparable to that of the true shape.

5.3 Application on real-world data

The final criteria to compare the approaches is their performance on real-world data. In our case, this is 2D monocular images of basking sharks in feeding motion. Annotated 2D landmarks are placed on the image and the pre-defined training shapes are used to estimate the 3D shape of the basking shark head skeleton. The estimated shape can only be assessed visually, as the ground truth is not available in this case. Figure 26 shows the 2D images with annotated landmarks and Figures 27, 28 and 29 show the estimated 3D shapes.

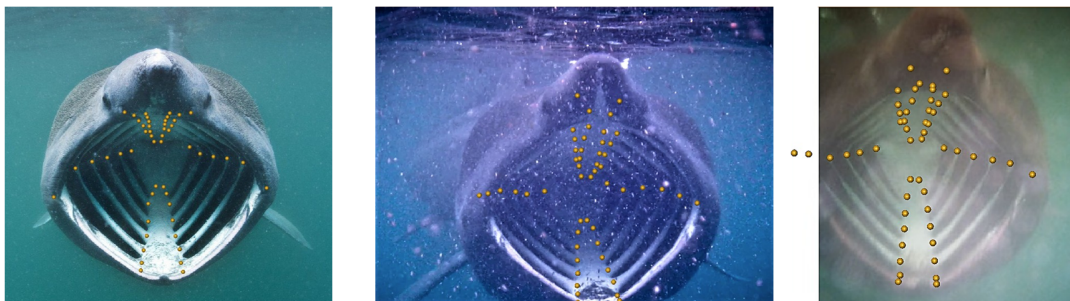
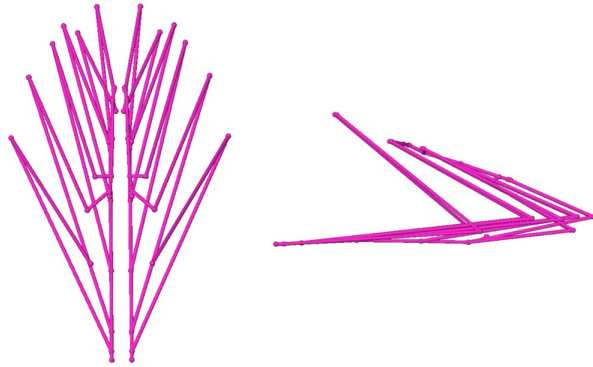
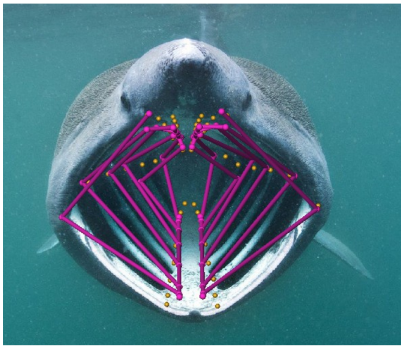
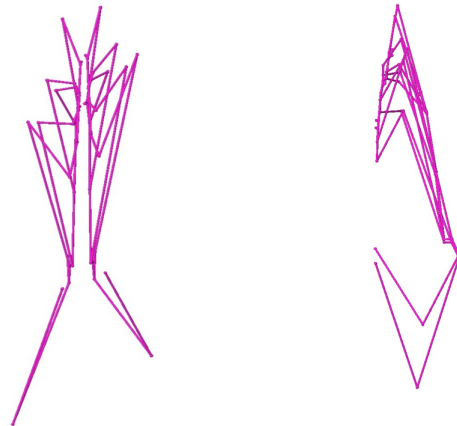
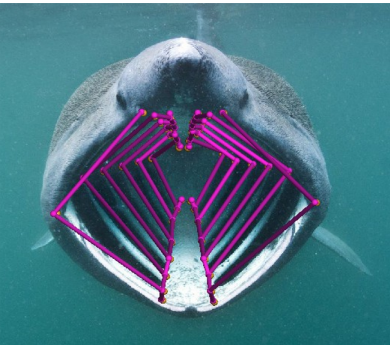


Figure 26: Images of basking sharks with annotated 2D landmarks (marked in yellow).

ASM non-convex



ASM convex



KSS

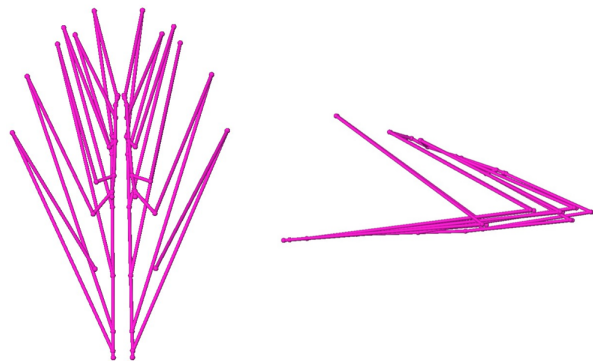
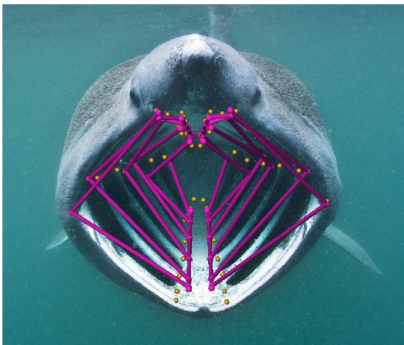
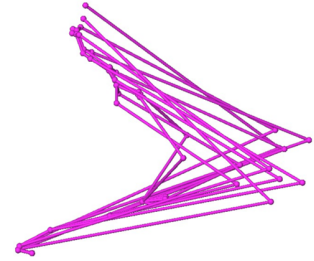
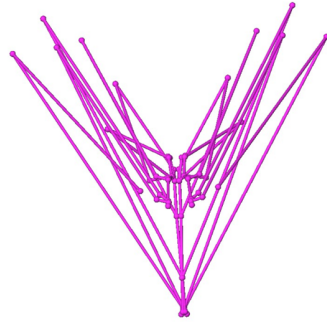
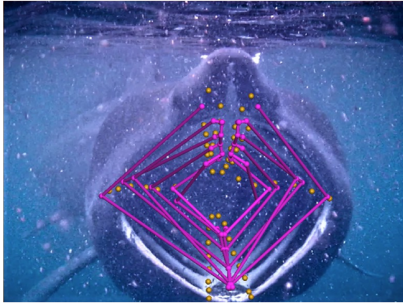
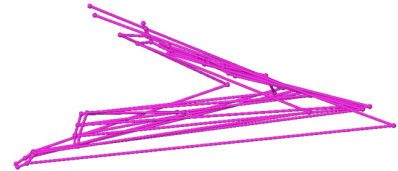
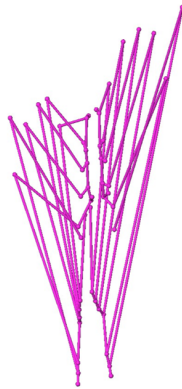
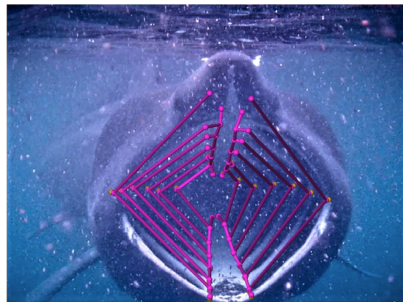


Figure 27: Estimated 3D shape of the head skeleton of a basking shark, from a single 2D image. Left to right: the estimated 3D shape in magenta , top view of the 3D shape and side view of the 3D shape.

ASM non-convex



ASM convex



KSS

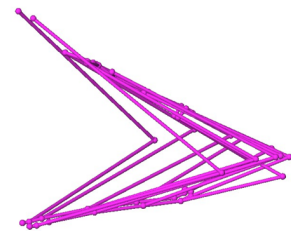
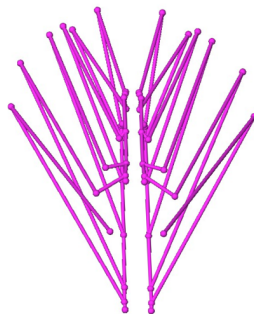
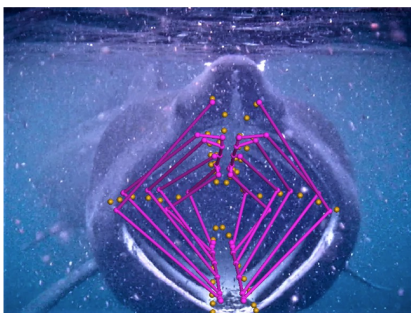
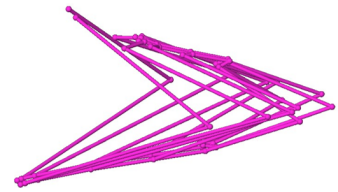
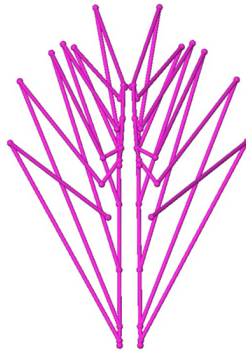
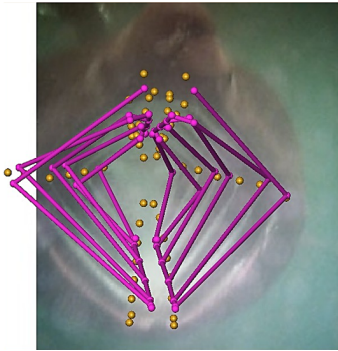
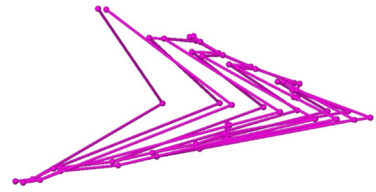
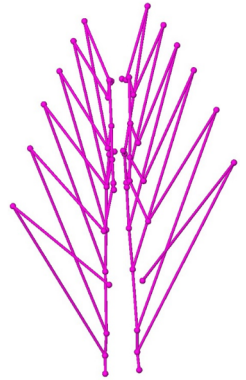
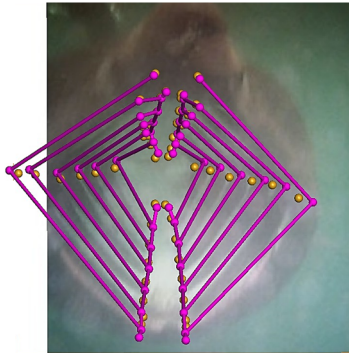


Figure 28: Estimated 3D shape of the head skeleton of a basking shark, from a single 2D image. Left to right: the estimated 3D shape in magenta , top view of the 3D shape and side view of the 3D shape.

ASM non-convex



ASM convex



KSS

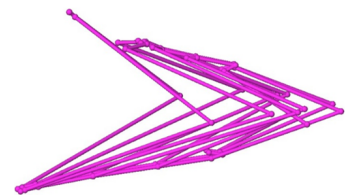
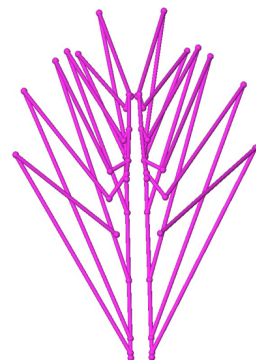
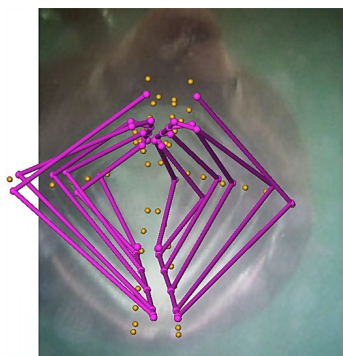


Figure 29: Estimated 3D shape of the head skeleton of a basking shark, from a single 2D image. Left to right: the estimated 3D shape in magenta, top view of the 3D shape and side view of the 3D shape.

We make the following observations from the three real-world applications:

- The estimated 3D shape in case of the ASM convex approach fits well with the 2D landmarks, but can appear very different from the training shapes. This can clearly be seen in Figures 27 and 28. Moreover, the estimated shape is sometimes squished or elongated, and requires manual scaling of the depth.
- For the ASM non-convex approach, the estimated 3D shape does not fit perfectly to the 2D landmarks. It performs well in some cases (Figures 27 and 29), and the estimated shape appears to be a plausible 3D representation of the head skeleton of the shark in the image, but not so well in others (Figure 28).
- KSS approach performs remarkably well in all three tested cases. Similar to the ASM non-convex approach, although the estimated 3D shape does not fit perfectly to the 2D landmarks, it appears to be a plausible representation.
- In Figures 27 and 29, the results from ASM non-convex and KSS approaches appear quite similar, but are indeed different. The shape distances between the estimated shapes are 0.056 and 0.002, respectively.
- The basking sharks in all three images are in feeding position with a fully open mouth, which causes the gill arches to “flare”. This flaring cannot be seen in the estimated shapes for any of the approaches. In all the cases, the gill arches are folded backwards, similar to when in closed mouth position.

6 Discussion and Outlook

This thesis was performed in association with the HFSP project titled “Integrating Materials, Behaviour, Robotics and Architecture in Giant Filter-Feeding Sharks”, which aims to study the filtering process of basking sharks, as a potential bio-inspiration for high-throughput dynamic filters, as well as from an evolutionary biological standpoint.

6.1 Comparing the approaches

The first approach used for the 2D-to-3D fitting problem, was the active shape model (ASM) approach, which estimates an unknown 3D shape as a linear combination of pre-defined training shapes. Two different formulations of this approach were tested, a non-convex and a convex one. The former is solved via an alternating minimization method and works well for the problem at hand. It is, however, sensitive to initialization and is unable to reliably estimate the unknown 3D shape, as seen in the case of bi-laterally asymmetric projections (Figure 22) and some real-world applications (Figure 28). The convex formulation is not sensitive to initialization and solves for a global optimum, however, the estimated 3D shapes often appear very different from the training shapes. Moreover, the estimated shapes can sometimes be bi-laterally asymmetric and even elongated or squished, needing a manual scaling of the depth. This is due to the fact that the training shapes are not aligned and can be rotated before interpolation, and the independent rotations and weights are estimated such that the objective in (12) is minimized. Thus, the similarity of the estimated shape to the training shapes is not guaranteed, which makes the convex formulation of the ASM approach is unattractive for the problem at hand.

A notable contribution of this thesis, is the development of a novel Kendall’s shape space (KSS) approach for the 2D-to-3D fitting problem. In contrast to the ASM approach, this approach uses a mathematical notion of “shape”, and performs computations on the Riemannian manifold representing the set of shapes, called Kendall’s shape space. It estimates the unknown 3D shape as a weighted Fréchet mean of some pre-defined training shapes, lying on this space. Similar to the ASM non-convex approach, an alternating minimization is used to solve for the weights and viewpoint parameters, in the optimization problem. In our experiments, this approach out-performed the ASM approach in almost all cases. This is due to a better interpolation between the training shapes, offered by the weighted Fréchet mean. The only drawback of KSS approach that we could find is longer computation time, compared to the other approaches. On average, it takes 1 hour to perform 50 iterations, compared to 7 minutes for the ASM non-convex and < 1 second (until convergence) for the convex approaches.

When testing the approaches on real-world images of basking sharks in Section 5.3, we observed that the 3D shapes estimated by the ASM non-convex and KSS approaches do not fit perfectly well with the 2D landmarks. Note that both the approaches assume the 2D landmarks to be the projection of a 3D shape,

using the weak-perspective camera, which is not the camera used for capturing the images. Moreover, an important pre-requisite for using them, is the availability of a comprehensive set of pre-defined training shapes, of the object being estimated. These might be the contributing reasons for this observation.

6.2 Basking shark head skeleton data

The head skeletons of basking sharks, which were segmented from the CT scans, were corrected by positioning the “spine” and the ventral plate to an anatomically plausible position. This was done by solving optimization problems which move the nodes of the shark skeleton, while preserving the lengths of its segments. The same method was used to open the mouth of the shark by positioning the “spine” higher and the ventral plate lower. This enabled us to successfully create pre-defined 3D training shapes of the head skeleton of basking sharks, to be used in the 2D-to-3D shape fitting problem. The limited research available on the relative movement of the different sub-regions of the branchial region (especially the gill arches) of this shark, restricted the accuracy of the prepared training shapes. We did, however, include all possible configurations of the skeleton, which preserved the length of the regions, while moving them in a plausible way. Rotating the gills and the pharyngobranchials is one such example. When fit to a set of 2D annotated landmarks, the output shape is interpolated from the training shapes. Hence, including a few extra shapes is not an issue, as long as the estimated 3D shape fits well to the 2D landmarks. This may not be the case with non-rigid objects with a larger variety of movement, for example, a human stick figure. In this case, it is maybe useful to use a sparse representation of the training shapes [ZLHD15, RKS12].

The creation of the piece-wise linear skeletons, using the 3D segmented regions from the CT scans, is in fact a reversible process. Sean Hanna and his team at Bartlett School of Architecture, UCL, used the symmetric piece-wise linear skeletons created by us to obtain the 3D segmented regions corresponding to it. Figure 30 visualizes this.

6.3 Outlook

Both the approaches studied for the 2D-to-3D shape fitting problem use a weak-perspective camera model, which assumes the depth of the object to be very small compared to the distance between the object and the camera. This simplifies the problem but is an inaccurate assumption for estimating the 3D shape of the head skeleton of a basking shark, as most images in their feeding position are captured by cameras close to the swimming sharks. Hence, in our case, the depth of the object is not small compared to the distance between the object and the camera. A more appropriate camera model would greatly improve the results, especially when estimating the 3D shape from real-world images of basking sharks. We believe this could be implemented for the ASM non-convex and KSS approaches. The ASM convex approach is derived using the assumption of a weak-perspective

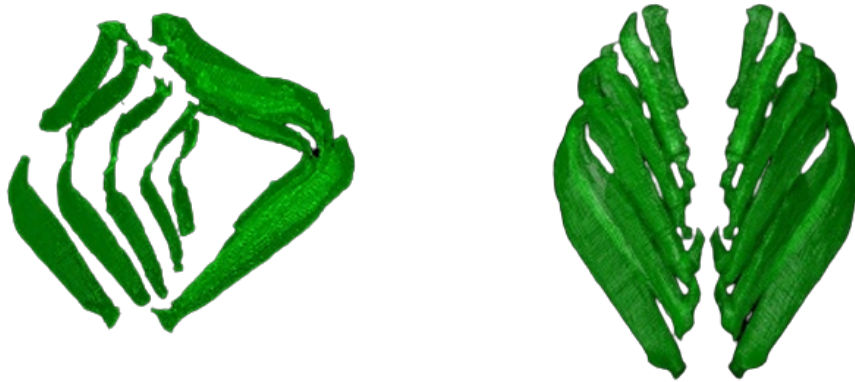


Figure 30: 3D head skeleton of a basking shark obtained from its piece-wise linear skeleton, in (left) open and (right) closed mouth positions. The open mouth position is viewed from the front while the closed mouth is viewed from the top (courtesy of Aurora Tairan Li and Sean Hanna)

camera, hence, it might be harder to change the camera model in this approach.

As acknowledged in the previous sections, due to the limited time provided for this thesis and the lack of data, these shapes were not anatomically accurate. Moreover, the gills were “flared” by rotation, only for a fraction of the pre-defined shapes. This favoured the estimated 3D shapes to have the gills folded backwards, i.e., not flared, even in open mouth positions (see Figures 27, 28, 29). A useful next step would therefore be the creation of more pre-defined shapes, especially in open mouth positions, for improved estimation.

As observed in Section 5, the ASM non-convex approach performed significantly better in the case of bi-laterally symmetric 2D projections and noisy 2D landmarks, compared to asymmetric projections. The reasons for this could be investigated further. Kendall’s shape space approach, on the other hand, performs consistently well for all the cases.

We opened the mouth of the basking shark head skeletons, by solving optimization problems which try to preserve the length of the linear segments of the skeleton while moving its nodes to new positions. Another way to do this could be by moving the nodes of the skeleton to new positions, such that the volume of the convex hull of the nodes is maximized and the lengths of the segments are preserved.

By estimating the 3D shape of the head skeleton using different frames in a 2D video of a feeding basking shark, the process can be animated in 3D. The animation of the skeleton can then be used to animate real-world volumetric data of basking sharks, similar to Figure 30. In Kendall’s shape space, given the initial, final and some intermediate positions of the head skeleton, a continuous

trajectory could be computed using geodesic curves on the space. To successfully do this, however, faster methods are needed. Speeding up KSS approach would be an appropriate option. It would also be interesting to study the relationship between the computation time of the approaches, and the number of training shapes.

During the course of this project, valuable data in form of surface scans and close-up diver footages of basking sharks were provided, which could not be used. A useful next step could also be using this data to attain kinematic information about the shark and incorporate this information in the current model.

References

- [AACM14] Antonio Agudo, Lourdes Agapito, Begona Calvo, and Jose MM Montiel. Good vibrations: A modal analysis approach for sequential non-rigid structure from motion. In *Proceedings of the IEEE Conference on computer vision and pattern recognition*, pages 1558–1565, 2014.
- [AB15] Ijaz Akhter and Michael J Black. Pose-conditioned joint angle limits for 3d human pose reconstruction. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1446–1455, 2015.
- [Alo90] John Y Aloimonos. Perspective approximations. *Image and Vision Computing*, 8(3):179–192, 1990.
- [ASS15] Boulbaba Ben Amor, Jingyong Su, and Anuj Srivastava. Action recognition using rate-invariant analysis of skeletal shape trajectories. *IEEE transactions on pattern analysis and machine intelligence*, 38(1):1–13, 2015.
- [BHB00] Christoph Bregler, Aaron Hertzmann, and Henning Biermann. Recovering non-rigid 3d shape from image streams. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, volume 2, pages 690–696. IEEE, 2000.
- [BL01] Andrea Bottino and Aldo Laurentini. A silhouette based technique for the reconstruction of human movement. *Computer Vision and Image Understanding*, 83(1):79–95, 2001.
- [BMAS14] Nicolas Boumal, Bamdev Mishra, P-A Absil, and Rodolphe Sepulchre. Manopt, a matlab toolbox for optimization on manifolds. *The Journal of Machine Learning Research*, 15(1):1455–1459, 2014.
- [Boo97] Fred L Bookstein. *Morphometric tools for landmark data*. 1997.
- [BPC11] Stephen Boyd, Neal Parikh, and Eric Chu. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.
- [BW91] Avinoam Beinglass and Haim J Wolfson. Articulated object recognition, or: How to generalize the generalized hough transform. In *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 461–462. IEEE Computer Society, 1991.
- [BWHK07] Anders Brun, Carl-Fredrik Westin, Magnus Herberthson, and Hans Knutsson. Intrinsic and extrinsic means on the circle-a maximum likelihood interpretation. In *2007 IEEE International Conference*

on Acoustics, Speech and Signal Processing-ICASSP'07, volume 3, pages III–1053. IEEE, 2007.

- [C⁺47] Augustin Cauchy et al. Méthode générale pour la résolution des systemes d'équations simultanées. *Comp. Rend. Sci. Paris*, 25(1847):536–538, 1847.
- [CBK03] KMG Cheung, Simon Baker, and Takeo Kanade. Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 1, pages I–I. IEEE, 2003.
- [CBMV18] Rudrasis Chakraborty, Jose Bouza, Jonathan Manton, and Baba C Vemuri. Manifoldnet: A deep network framework for manifold-valued data. *arXiv preprint arXiv:1809.06211*, 2018.
- [CHCF18] Karly E Cohen, L Patricia Hernandez, Callie H Crawford, and Brooke E Flammang. Channeling vorticity: modeling the filter-feeding mechanism in silver carp using μ ct and 3d piv. *Journal of Experimental Biology*, 221(19):jeb183350, 2018.
- [CTCG95] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.
- [CWLZ13] Chen Cao, Yanlin Weng, Stephen Lin, and Kun Zhou. 3d shape regression for real-time facial animation. *ACM Transactions on Graphics (TOG)*, 32(4):1–10, 2013.
- [FDC⁺21] Rasha Friji, Hassen Drira, Faten Chaieb, Hamza Kchok, and Sebastian Kurtek. Geometric deep neural network using rigid and non-rigid transformations for human action recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12611–12620, 2021.
- [Gow75] John C Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.
- [HBA20] Nadia Hosni and Boulbaba Ben Amor. A geometric convnet on 3d shape manifold for gait recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 852–853, 2020.
- [HH14] Stephan Huckemann and Thomas Hotz. On means and their asymptotics: circles and shape spaces. *Journal of mathematical imaging and vision*, 50(1):98–106, 2014.
- [How04] Nicholas R Howe. Silhouette lookup for automatic pose tracking. In *2004 Conference on Computer Vision and Pattern Recognition Workshop*, pages 15–22. IEEE, 2004.

- [HR12] Mohsen Hejrati and Deva Ramanan. Analyzing 3d objects in cluttered images. *Advances in Neural Information Processing Systems*, 25, 2012.
- [HT92] Andrew Hill and Christopher J Taylor. Model-based image interpretation using genetic algorithms. *Image and Vision Computing*, 10(5):295–300, 1992.
- [HWR⁺91] Geoffrey E Hinton, Christopher KI Williams, Michael D Revow, et al. Adaptive elastic models for hand-printed character recognition. In *NIPS*, volume 4, pages 512–519, 1991.
- [JLT⁺12] Hanqing Jiang, Haomin Liu, Ping Tan, Guofeng Zhang, and Hujun Bao. 3d reconstruction of dynamic scenes with multiple handheld cameras. In *European Conference on Computer Vision*, pages 601–615. Springer, 2012.
- [Ken84] David G Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London mathematical society*, 16(2):81–121, 1984.
- [Kli20] Christian Peter Klingenberg. Walking on kendall’s shape space: Understanding shape spaces and their coordinate systems. *Evolutionary Biology*, 47(4):334–352, 2020.
- [KWT88] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.
- [LMPF07] E Scott Larsen, Philippos Mordohai, Marc Pollefeys, and Henry Fuchs. Temporally consistent reconstruction from multiple video streams using enhanced belief propagation. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007.
- [LYO⁺90] P Lipson, Alan L Yuille, D O’keeffe, J Cavanaugh, J Taaffe, and D Rosenthal. Deformable templates for feature extraction from medical images. In *European Conference on Computer Vision*, pages 413–417. Springer, 1990.
- [MDA15] Dougal Maclaurin, David Duvenaud, and Ryan P Adams. Autograd: Effortless gradients in numpy. In *ICML 2015 AutoML workshop*, volume 238, page 5, 2015.
- [MGLB⁺20] Nina Miolane, Nicolas Guigui, Alice Le Brigant, Johan Mathe, Benjamin Hou, Yann Thanwerdas, Stefan Heyder, Olivier Peltre, Niklas Koep, Hadi Zaatiti, et al. Geomstats: A python package for riemannian geometry in machine learning. *Journal of Machine Learning Research*, 21(223):1–9, 2020.

- [MKGH15] Armin Mustafa, Hansung Kim, Jean-Yves Guillemaut, and Adrian Hilton. General dynamic scene reconstruction from multiple view video. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 900–908, 2015.
- [MKW91] KV Mardia, JT Kent, and AN Walder. Statistical shape models in image analysis. In *Proceedings of the 23rd Symposium on the Interface, Seattle*, pages 550–557, 1991.
- [MM06] Greg Mori and Jitendra Malik. Recovering 3d human body configurations using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(7):1052–1062, 2006.
- [MPTS14] EW Misty Paig-Tran and AP Summers. Comparison of the structure and composition of the branchial filters in suspension feeding elasmobranchs. *The Anatomical Record*, 297(4):701–715, 2014.
- [MVGv09] Theo Moons, Luc Van Gool, and Maarten Vergauwen. *3D reconstruction from multiple images: principles*. Now Publishers Inc, 2009.
- [NA93] Chahab Nastar and Nicholas Ayache. Fast segmentation, tracking, and analysis of deformable objects. In *1993 (4th) International Conference on Computer Vision*, pages 275–279. IEEE, 1993.
- [NYHSvT20] Esfandiar Nava-Yazdani, Hans-Christian Hege, Timothy John Sullivan, and Christoph von Tycowicz. Geodesic analysis in kendall’s shape space with epidemiological applications. *Journal of Mathematical Imaging and Vision*, 62(4):549–559, 2020.
- [NYHvT19] Esfandiar Nava-Yazdani, Hans-Christian Hege, and Christoph von Tycowicz. A geodesic mixed effects model in kendall’s shape space. In *Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy*, pages 209–218. Springer, 2019.
- [Pen06] Xavier Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.
- [PF01] Ralf Plänkers and Pascal Fua. Tracking and modeling people in video sequences. *Computer Vision and Image Understanding*, 81(3):285–302, 2001.
- [RCA⁺13] Alex D Rygg, Jonathan PL Cox, Richard Abel, Andrew G Webb, Nadine B Smith, and Brent A Craven. A computational study of the hydrodynamics in the nasal region of a hammer-head shark (*sphyrna tudes*): implications for olfaction. *PLoS One*, 8(3):e59783, 2013.

- [RKS12] Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Reconstructing 3d human pose from 2d image landmarks. In *European conference on computer vision*, pages 573–586. Springer, 2012.
- [Rod40] O Rodriguez. Des lois geometriques qui regissent les desplacements d’un systeme solide dans l’espace et de la variation des coordonnees provenant de déplacements consideres independamment des causes qui peuvent les produire. *J. Mathematiques Pures Appliquees*, 5:380–440, 1840.
- [Sim08] David W Sims. Sieving a living: a review of the biology, ecology and conservation status of the plankton-feeding basking shark *cetorhinus maximus*. *Advances in marine biology*, 54:171–220, 2008.
- [SNP16] Marta Sanzari, Valsamis Ntouskos, and Fiora Pirri. Bayesian image based 3d pose estimation. In *European conference on computer vision*, pages 566–582. Springer, 2016.
- [SWH⁺05] Detlev Stalling, Malte Westerhoff, Hans-Christian Hege, et al. Amira: A highly interactive system for visual data analysis. *The visualization handbook*, 38:749–67, 2005.
- [YHC92] Alan L Yuille, Peter W Hallinan, and David S Cohen. Feature extraction from faces using deformable templates. *International journal of computer vision*, 8(2):99–111, 1992.
- [ZLHD15] Xiaowei Zhou, Spyridon Leonardos, Xiaoyan Hu, and Kostas Daniilidis. 3D shape estimation from 2D landmarks: A convex relaxation approach. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 4447–4455, 2015.