A New Landmark-Independent Tool for Quantifying and Characterizing Morphologic Variation

Sara Rolfe

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Linda Shapiro, Chair

Timothy Cox

Jenq-Neng Hwang

Program Authorized to Offer Degree:
Electrical Engineering

University of Washington

Abstract

A New Landmark-Independent Tool for Quantifying and Characterizing Morphologic Variation

Sara Rolfe

Chair of the Supervisory Committee:
Title of Chair Linda Shapiro
Department of Chair

As advances in multi-dimensional imaging are made, there is a growing need for tools that enable quantitative assessment of both embryonic and postnatal structure and form. To this end, this dissertation develops a landmark-independent, deformable-registration based framework for quantifying and characterizing morphologic variation. This framework can utilize 3D surface images generated by any multi-dimensional biomedical imaging modality and provides quantitative assessment of potentially biologically relevant changes in 3D shape whether natural, caused by mutation, or influenced by non-genetic factors. In this work, using different example datasets, the utility and sensitivity of the tools developed in this work are demonstrated: 1) to detect subtle morphological phenotypes in specimens; 2) to quantify morphologic change over a normal developmental time series of embryonic development; and 3) to determine the impact of transient localized genetic manipulation and environmental change on morphogenesis. The development of these tools will assist in defining trajectories of normal growth and help underpin hypotheses as to the developmental basis and the specific role of the causative gene in conditions causing disruptions to normal growth.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

DEDICATION

To my family,

my parents for the opportunity and encouragement to pursue my dream

and my beloved husband Justin for his support in making it a reality.

Chapter 1

INTRODUCTION

1.1 Motivation

Traditionally, embryologic and developmental biology studies have relied on subjective assessments of gross phenotypic changes, whether for description of normal changes during embryogenesis or how mutants differ from controls. However, as advances in multi-dimensional imaging are made, there will be a growing need for tools that enable quantitative assessment of both embryonic and postnatal structure and form. The development of these tools will assist in defining trajectories of normal growth and help underpin hypotheses as to the developmental basis and the specific role of the causative gene or the contributions of in utero environmental factors in conditions causing disruptions to normal growth. Specific examples of the study of such conditions is the work being done at Seattle Children's Research Institute (SCRI) on craniofacial microsomia and cleft lip and/or palate.

1.1.1 Craniofacial microsomia

Human faces are almost never perfectly symmetric. However, in some individuals the asymmetry can be quite marked, which is clinically referred to as craniofacial microsomia. The condition usually involves underdevelopment, or hypoplasia, of the midface (upper jaw and nose), the lower jaw (or mandible), as well as positioning and/or malformation of the ear. Craniofacial microsomia is one of the most common facial birth defects, occurring in approximately 1 in 3500 live births [21]. The cause of this relatively common anomaly remains unknown, although both genetic and non-genetic factors are believed to contribute to the presentation. In the rarer familial cases, there is no consistency in the side of the face most affected. A detailed analysis of differences between the bilaterally paired structures would assist in understanding the pathogenetic mechanisms of this

condition and help elucidate factors that impact the severity of the presentation.

1.1.2   Cleft Lip and/or Palate

Clefting of the lip and/or palate (CL/P) is a common human birth defect that occurs when the lip or lip and palate fail to fuse during the facial development. The consequences for the infant are a gap or space in the lip or lip and palate which can lead to difficulties with necessary functions such as breathing, feeding, and speech development. CL/P is the most common facial birth defect, occuring in approximately 1 in 700 live births [22]. In spite of the prevalence of this defect, its genesis is still not fully understood. Study of this defect is complicated by the fact that it is caused by a combination of genetic and environmental factors with complex interactions. Facial development involves the coordinated growth of multiple three-dimensional structures, which grow towards each other in pairs and fuse into a single craniofacial structure [29]. Any disruption of these structures or the timing of their growth can result in the failure of these embryonic tissues to meet and fuse. Some of the factors that have been identified as potential influences are: small facial prominences, slow or uncoordinated growth of prominences, failure of prominences to fuse when in contact, and any environmental effects which exacerbate these factors [38]. A detailed examination of these developmental factors would provide valuable information on which processes play a major role in clefting.

1.2   Problem Statement

Quantitative assessment is complicated by the difficulty of analyzing three-dimensional shape change. Traditional shape analysis methods generally rely on the use of landmarks manually placed on an image by an expert [39], [40]. These landmarks are chosen at locations that are easy to identify across specimens, such as at the juncture between structures. Once identified, the relative difference in position of the landmarks can be compared across individuals to provide important shape information. However, the use of landmarks has several drawbacks. Manual placement of landmark points for shape analysis is tedious and subject to variability. In addition, no shape information is provided about regions between landmark points, which may be scientifically interesting. For some biological structures, such as embryos, very few landmarks are available that can be reliably identi-

fied across samples of different ages because of the significant morphological and growth changes that occur during embryogenesis.

Motivated by collaborations with researchers at SCRI, this work develops a landmark-independent deformable registration-based framework for the analysis of 3D images generated by any multi-dimensional imaging modality. This framework is used to tackle the three following problems:

1. detecting subtle morphological phenotypes in specimens, in this case mandibles from control and mutant mice at a single postnatal time-point (acquired using micro-computed tomography);

2. quantifying morphologic change over a normal developmental time series of embryonic facial development acquired using optical projection tomography; and

3. determining the impact of transient localized genetic manipulation and environmental (or epigenetic) change on facial morphogenesis.

1.3 Paper Outline

In Chapter 2, literature related to the method developed in this work is reviewed. This includes work done in geometric morphometrics, deformation-based morphology, and symmetry evaluation of biological samples. The base methods for generating the deformable registration vector field features used throughout this paper are introduced in Chapter 3. These features are validated using synthetic data in Chapter 4. Chapter 5 introduces a symmetry analysis tool which makes use of the deformable field vector features to assess asymmetry in bilaterally paired structures. This tool is demonstrated on the hemi-mandibles of control mice and mice expressing craniofacial microsomia due to a single gene defect. Chapter 6 introduces a framework for summarizing, integrating, and describing the spatial distribution of the image difference features and defines a distance metric to quantify similarity between shape change trajectories. Chapter 7 applies this framework to quantify morphologic change over a normal developmental time series of chick embryos, and in Chapter 8 it is used to analyze chick embryos that have received a transient genetic manipulation resulting in an orofacial cleft. Finally, the conclusion in Chapter 9 will summarize the contributions of this work and suggest possible future applications and directions.

Chapter 2

RELATED LITERATURE

2.1 Geometric Morphometric methods

Geometric morphometrics (GM) is a powerful tool for analyzing shape and size. It is the most commonly used method for measuring the difference in shape between two biological samples. This method relies on the use of landmarks which can be precisely identified on each individual and are usually placed manually by an expert. Landmarks are typically located at the junctures of identifiable structures or along the object outline at points of maximum curvature. Once the landmarks have been identified on all samples, their relative position can be compared across groups. The gold standard for aligning landmarks and obtaining their shape information is the Procrustes method [12], [33]. This method finds an affine transformation that minimizes the distance between shapes, which is defined as the square root of the sum of squared distances between homologous landmarks. This transformation removes position and orientation information. Scale information can optionally be removed at this step as well. Comparisons between aligned sets of landmarks can then be made across groups using statistical analysis [39], [40].

GM methods were applied by Hallgrímsson *et al.* to study craniofacial shape in [13], where statistics capturing the shape and size variance of cleft-liable and normal strains of mice were compared. This method was extended to the study of embryonic development by Young *et al.* in [38], which used GM methods to discriminate between cleft-liable and normal strains of mice during development. The results suggested that the statistics had discriminating power but were limited by the fact that the landmarks were chosen on two two-dimensional planes, reducing the shape information they provided. This technique was improved upon by Parsons *et al.* [29], who demonstrated three-dimensional GM analysis of mice embryos.

The use of landmarks poses several challenges. Manual placement of the landmarks on the image is time-consuming and error-prone. Some mitigating steps have been taken to reduce the measurement error. Hallgrímsson *et al.* [13] used landmarks that were digitized twice to try to re-

duce error. Zelditch *et al.* [39, 40] reflected landmarks across the midline of the skull and averaged bilaterally homologous landmarks, though this precluded symmetry analysis. The use of landmarks also poses a challenge when comparing groups where a disease or species difference causes significant structural variation between the groups. Landmarks identified on one group may be difficult or impossible to locate on the other. In addition, landmark data can be unavailable over large regions of scientifically interesting areas in an image. This is especially true for studies of embryonic development such as those done by Young *et al.* [38] and Parsons *et al.* [29], since embryos have few structures that can be reliably identified across samples.

2.2 Adapted Geometric-Morphometric methods

One approach to solving the problem of insufficient landmark data is to supplement landmarks with additional points on the object surface, called semi-landmarks. Polly [32] conducted a study on the ankle bones of eleven species of carnivorous mammals. Semi-landmarks were created by placing a cylindrical grid of points around the bone and projecting this grid onto the surface of the bone. These semi-landmarks were then used for Procrustes analysis. A second method, described by Bookstein [3], used thin-plate splines to assign semi-landmarks on curving outlines of biomedical images.

A second approach to solving the problem of insufficient landmark data is to eliminate their use. Kristensen *et al.* [17] introduced a method to describe differences between mouse skulls without landmarks. Images were first aligned using a rigid registration and an average sample shape is determined. Local shape differences between the average image and individual images were determined by a surface-to-surface distance measure on the superimposed images. However, since the surface-to-surface measure does not guarantee that the point pairs compared are homologous, this method could not be used to characterize the shape change detected. It was presented as a screening tool to be used alongside established morphometric methods.

2.3 Deformation-Based Morphology

In the human brain, reliable landmarks are difficult to identify due to the lack of structural features and individual variability. This has motivated the development of alternative morphology techniques. Ashburner *et al.* [1] introduced the term deformation-based morphology to describe a

method for analyzing group differences in brain shape using deformation vector fields, which arose out of a need to address the problem of brain registration. The deformation vector field of each individual described the spatial normalization of that image to a standard anatomical space. In effect, the deformation vector field provided continuous landmark data, since each point in the individual image was mapped to the corresponding point in the reference space. Once the deformation vector fields were calculated for each individual, Ashburner [1] described the application of multivariate statistics to the vector field parameters that were similar to those used in traditional GM analysis. These statistics were used to discriminate between groups with different gross morphometry, such as male and female subjects. The statistical analysis of the vector fields was extended by Cao and Worsley in [5], where significant differences in deformations were detected between two groups using high values of Hotelling's T^2 statistics, which indicated a change in mean of the vector deformations at that point in the object.

Wollney *et al.* presented an alternative approach to the interpretation of these vector fields in [37], where the inherent topology of the deformation vector field was used to locate regions of significant change rather than group statistics. In this study, two brain images from a patient suffering from Alzheimer disease were registered to each other, producing a deformable vector field describing brain shape change between the scans. The critical points in the vector field were identified, and those classified as attractors or repellors were interpreted as areas of tissue growth or loss. Thirion and Calmon [36] applied these methods to a time-series of images of a patient with Multiple Sclerosis to investigate brain volume change over time.

Deformation morphology has also been applied to more standard morphometric measurement applications. Ólafsdóttir *et al.* presented a computational mouse atlas for the skull in [25] that represented the average of a group of normal, wild-type mice. This atlas was warped to each subject from a normal or mutant strain that serves as a model of Crouzon syndrome to evaluate individual differences. In several following papers, Ólafsdóttir *et al. et al.* evaluated methods for interpreting and reducing the dimensionality of the dense field of shape differences. PCA was first used to find the modes of the deformable field that could discriminate between the normal and Crouzon mutation groups [24]. However, only the first model was capable of describing group differences. Independent Component Analysis (ICA) and Sparse Principal Component Analysis (SPCA) were then evaluated with the aim of distributing the discriminating variation over more modes [26]. ICA

and SPCA had similar performance and discovered up to six deformation modes. SPCA was recommended over ICA because of its simple and flexible implementation.

2.4 Symmetry Evaluation

Deformation morphology techniques have been useful in assessing the symmetry of a sample. Deformation-Based Asymmetry (DBA), a method for detecting asymmetry in the brain surface was introduced by Lancaster *et al.* [18]. In this study, the brain image of each subject was segmented and a separate image was created from each hemisphere. An average image from all subjects was created for each hemisphere. One average hemisphere image was mirrored and warped to the other to estimate the spatial differences between them. To quantify the asymmetry between average hemisphere images a statistical measure, called the d-value, was defined as the ratio of the displacement vector magnitude to the estimated within-group warping variance at that point. The proportion of d-values above a fixed level was used to test for asymmetry. Combes and Prima [7] computed an individual map of brain symmetry by deformably registering an image to itself mirrored over the line of symmetry. For group comparisons, all brain images were deformably registered to a perfectly symmetrical template. This template was created by choosing a reference image and mirroring it across its symmetry plane. Statistical tests were then performed at each point on the template to assess group symmetry.

A similar method was applied by Ólafsdóttir *et al.* [27] to assess craniofacial symmetry in a mouse model. A symmetric atlas was generated by mirroring an average image across the mid-saggital plane. The atlas was then registered to each subject in a group and the resulting deformation field was used to asses individual asymmetry. In doing so, an asymmetry measure was proposed and defined as the difference in magnitude of the deformation vectors at corresponding points on the left and right side of the symmetry plane. This measure gave a point-wise estimation of the asymmetry of an individual. Using this measure, mice from the aforementioned Crouzen mouse model were compared to a control group and were found to have a higher average level of asymmetry. This method was also applied to CT images of human mandibles in [10].

Chapter 3

BASE FRAMEWORK

The methods presented in this work utilize low-level, mid-level, and high-level features generated from comparisons between high resolution optical scans of biological specimens. This chapter will detail the generation of these surfaces and the extraction of the low and mid-level features.

3.1 Contour Generation

While high-resolution optical scans can produce detailed images of microscopic specimen, they also pose several challenges when used for shape evaluation. Noise in the image can cause bright pixels that are reflected outside the object, gaps or holes where the object is obscured, or indistinct borders between the object and the background. In addition, the internal structure of the sample, such as the brain cavity and eye lenses, can result in a complex segmentation rather than the external contour that is desired. For this application, a method based on geodesic active contours is used to address these issues and locate the surface of the sample. This method is designed to operate on 3D scan data, but is also capable of handling 2D images.

3.1.1 Geodesic Active Contours

The geodesic active contour method is based on the classical snakes segmentation framework [16]. The snakes approach associates a curve C in image I with an energy $E(C)$ defined by:

$$E(C) = \alpha \int_0^1 |C'(q)|^2 \, dq - \lambda \int_0^1 |\Delta I(C(q))| \, dq, \qquad (3.1)$$

where α and λ are real, positive constants. The first term in equation 3.1 represents the smoothness, or internal energy, of the contours detected. The second term represents the attraction force towards the object edges, or the external energy of the contour. The optimal contour placement is found when the energy E is minimized for given values of α and λ.

<div align="center">(a)   (b)   (c)</div>
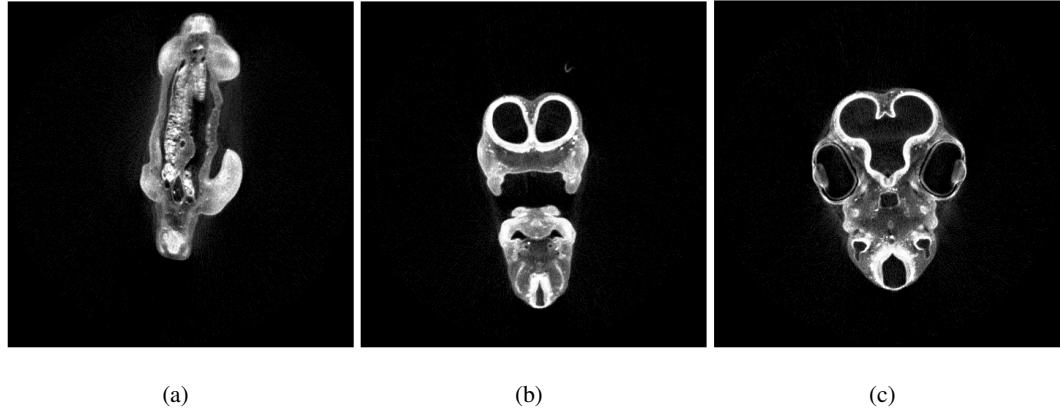
Figure 3.1: Three axial slices from a reconstruction of a 3D OPT image of a chick embryo

The functional in equation 3.1 depends on the parameterization of q, which is an undesirable quality. In [6], geodesic active contours are developed as an alternative geometric formulation of the snakes framework which is shown to be mathematically equivalent, but avoids the arbitrary parameterization of the curve. In this method, the curve is embedded as a zero level set of a higher dimensional surface. The entire surface is then evolved to minimize the energy functional. Geodesic active contours were shown to have more stable solutions and improved adherence to object edges when compared to methods using the classical snakes formulation.

3.1.2 ITK Implementation of Contour Extraction

Geodesic active contour segmentation of the tomographic scan images was implemented using the Insight Toolkit (ITK), which provides many general image processing filters that can be connected in a pipeline to accomplish application specific tasks[1]. The pipeline of filters used for this application is shown in figure 3.2.

The Geodesic Active Contour filter 3.2 requires two input images. The first is the level set required to initialize the contour, which should be relatively close to the desired segmentation output. This image is produced by the first branch of the pipeline in figure 3.2. The second input required is the feature image used to attract the contour to the edges of the object in the image. This image is

---

[1]ITK is freely available at `www.itk.org`
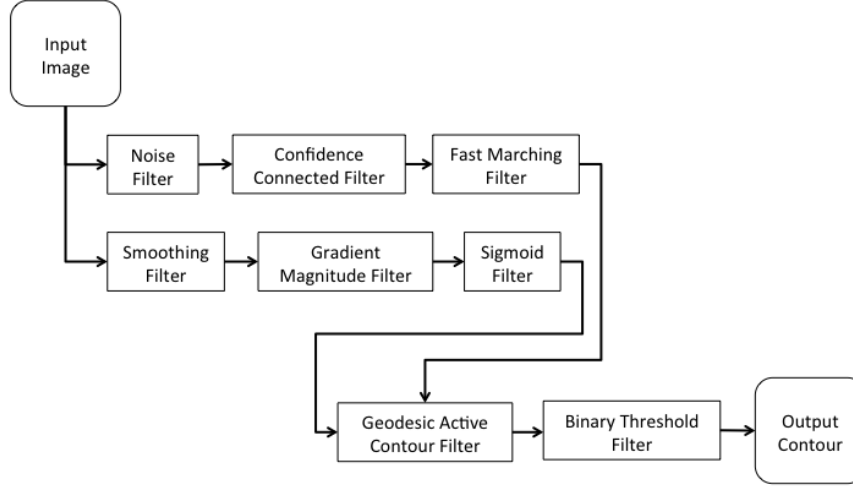
Figure 3.2: Filter pipeline for contour extraction from 3D scan data

produced by the second branch of the pipeline. In the following section, the individual filters in the pipeline are briefly described [15].

Noise Filter The Otsu threshold image filter is used to classify the image into a foreground cluster and a background cluster such that the within-class variance is minimized and the between-class variance is maximized [28]. Pixels assigned to the background cluster are set to a predefined output level. This filter reduces low-level noise in the background of the image.

Confidence Connected Image Filter The confidence connected image filter extracts a set of pixels from the image that are spatially connected and share similar statistics. The filter is initialized with connected seed pixels and the mean and standard deviation of these pixels are calculated. Pixels which are spatially connected to the seed pixels and whose values are similar are assigned to the region using the criteria:

$$I(X) \; \epsilon \; [m - f\sigma, m + f\sigma], \tag{3.2}$$

where I is the image, X is the position of a pixel being considered for inclusion, m and σ are the mean and standard deviation of the region intensities, and f is a factor describing the tolerance of the segmentation. This process is repeated until no new neighbors are added. This filter is used

to identify the background surrounding the object in the image. The interior of the object in the image may contain pixels with statistics similar to the background but when the exterior contour of the object is closed, these will not be included in the background segmentation. This eliminates the segmentation of additional structures within the exterior contour. The object in the image is represented by the complement of the pixels identified as background by the confidence connected image filter.

Fast Marching Image Filter The fast marching image filter computes the distance map for a given set of seed pixels. A distance map labels each pixel in an image with the distance to the nearest seed pixel. The pixels that were not identified as background by the confidence connected image filter are used as seed pixels. The output of this filter is an image in which the value at each pixel represents the distance from that pixel to the nearest pixel determined to be part of the object. This image is used as the initial level set for the geodesic active contour filter.

Smoothing Filter The curvature anisotropic diffusion filter is a smoothing filter that performs anisotropic diffusion. This method reduces image noise without removing significant parts of the image content, such as edges, lines or other details [30]. This filter is used to smooth the input image before calculating the image gradient.

Gradient Magnitude Image Filter The gradient magnitude recursive Gaussian image filter computes an approximation of the image gradient using a finite differences method. This filter outputs an image where the value of each pixel is the magnitude of the input image gradient at that location. The gradient magnitude is used to help detect edges in the input image.

Sigmoid Filter The sigmoid image filter maps a specific range of intensity values into a new intensity range. This focuses attention on a particular set of values and progressively attenuates the values outside that range. In this application, this filter is used to intensify the differences between regions of low and high values in the gradient magnitude image.

Geodesic Active Contour Filter The geodesic active contour level set image filter is initialized with the level set from the fast marching image filter and the feature image representing the image

(a)   (b)   (c)
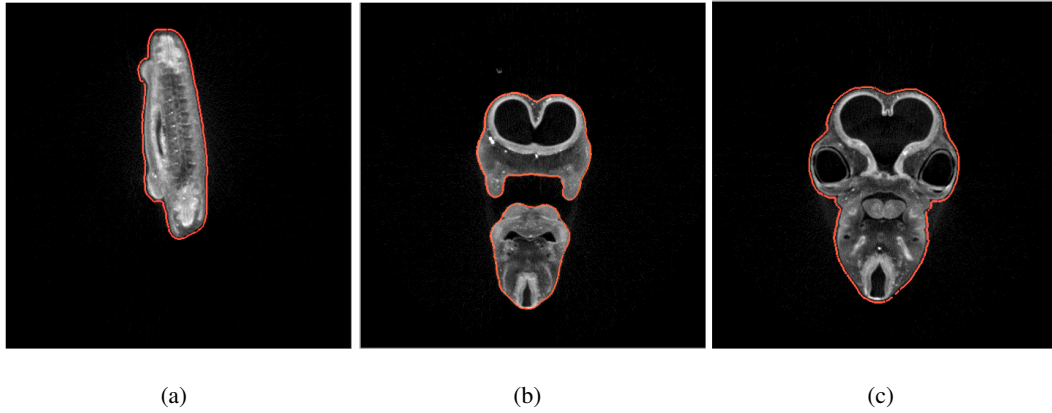
Figure 3.3: Three axial slices from a reconstruction of a 3D OPT image of a chick embryo with the exterior contour marked in red.

edges from the sigmoid image filter. The filter iteratively evolves the level set to minimize the energy of the functional in equation 3.1. The zero level set of the final surface represents the segmented image boundary.

Binary Threshold Filter   The binary threshold image filter is used to transform the output of the geodesic active contour filter into a binary image. The pixel values within the zero level set are set to 1 and all others are set to 0. The output of this threshold filter is the final object contour. Figure 3.3 shows an example of this contour output mapped back to the corresponding axial slices of the original image.

3.2   Initial Alignment

The final step in pre-processing the images before feature extraction is the alignment of the extracted surface contours. This eliminates differences in the images due to global position and orientation, while preserving local differences that are important for shape analysis. This step greatly influences the interpretation of shape differences between two contours, since it provides the reference frame within which the differences are quantified.

This initial alignment can be accomplished using one of two methods. The first is a blind, affine registration where two surfaces are aligned automatically so that the differences between the two

surfaces are minimized. The second method uses a small number of manually placed landmarks that allow an expert to specify a method for alignment that is specific to the object being analyzed.

3.2.1 Automatic Affine Registration

For a vector x, the affine transformation y is defined mathematically as:

$$y = Ax + b, \tag{3.3}$$

where A is a matrix holding parameters that describe rotation, scaling and shearing, and b is a vector holding the translation parameters [35]. The difference between two points in an image and its affine transform is linear. Applying this transform to the extracted contours rigidly registers them to the same space using only linear transformations and a translation. The optimal transformation is found by minimizing the summed square differences between the intensity values of the images.

3.2.2 Landmark Registration

Some applications require alignment that minimizes the distances in a specific region rather than over the entire shape. For example, when analyzing the mid-facial development of chick embryos, it is necessary to align the images with respect to the face, rather than aligning the entire heads. To accomplish this, a small number of landmark points are identified as crucial to the desired image alignment. The landmarks are used to align images so that the summed square distances between corresponding landmarks are minimized. Figure 3.4 shows an example of the five landmark points used for the initial alignment of the embryonic chick images with respect to the mid-face.

3.3 Feature Generation

Once two contours have been transformed into the same space, a deformable registration is used to assess local differences at every point. Low and mid-level features are extracted that will be used to characterize image regions and rank their importance.

3.3.1 Deformable Registration

The goal of an image registration is to determine the spatial transform mapping points from a source image to homologous points in a target image. The output of the registration is a dense deformation

(a) Frontal   (b) Lateral
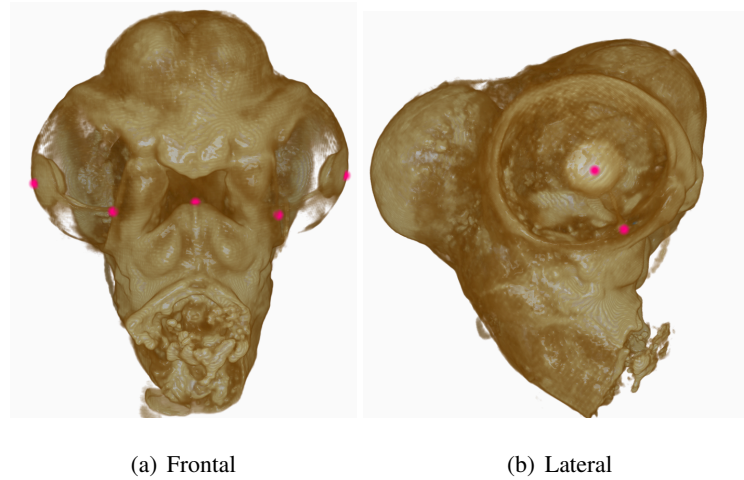
Figure 3.4: Landmark placement for initial alignment of chick embryo heads. Landmarks are shown in red.

vector field in which the vector at each point describes the spatial transformation of that point. When applied to three-dimensional images of two objects, these vectors reflect the structural differences between the source and target objects.

An image registration algorithm requires a transformation model which describes how the source image can be changed to match the target image, a metric to determine the fitness of the current registration iteration, an optimizer that adjusts the transform parameters to improve the metric, and an interpolator that applies the transform to the image. The flow chart in figure 3.5 shows how these components are interconnected. For this application the transform chosen is the B-spline deformable transform, the metric is mutual information, and the optimizer is limited memory Broyden-Fletcher-Goldfarb-Shannon minimization with simple bounds (LBFGSB). These components were implemented using ITK and the function of each is described briefly in the following sections [15].

B-spline Deformable Transform

The B-spline transform is a non-linear transformation that is widely used due to its general applicability and computational efficiency [9]. The deformation is modeled as a cubic grid of control points $\vec{\lambda_i}$ placed over the volume. The control points are moved to adjust the transform, which is defined by the deformation $\vec{g}(\lambda_i)$ at each control point. A B-spline interpolation kernel is used to interpolate
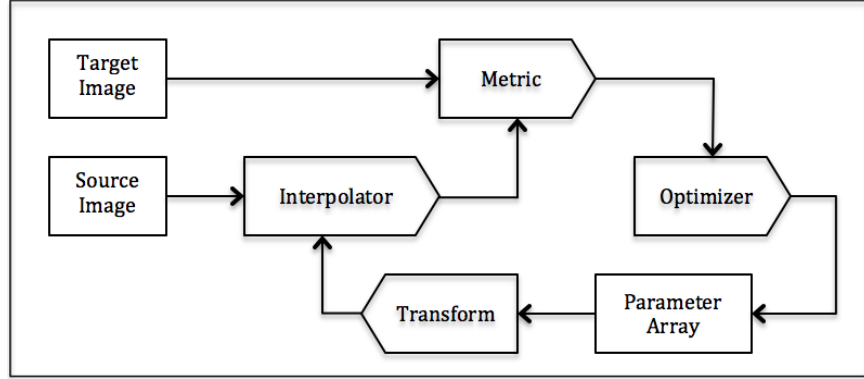
Figure 3.5: Flow chart of image registration

$\vec{g}(\lambda_i)$ to get the globally smooth deformation $D(\vec{x})$ which is defined at each point $\vec{x}$ in the image.

Mutual Information Metric

The similarity between the target and deformed source image is computed after each iteration of the B-spline transform to measure its performance. The metric chosen for this application is the mutual information metric, which uses mutual information between two images as the similarity criterion [20]. Mutual information I between an image A and an image B can be defined as

$$I(A, B) = H(B) - H(B|A), \tag{3.4}$$

where $H(B)$ is the Shannon entropy of image B, computed on the probability distribution of image values. $H(B|A)$ is the conditional entropy and is based on the conditional probabilities $p(b|a)$, the chance that a voxel in image B will have value b given that the corresponding voxel in A has value a. Since Shannon entropy can be interpreted as a measure of uncertainty, the mutual information reflects how much the uncertainty of image B is decreased when the information in image A is known. When the two images are aligned, the amount of information they contain about each each other is maximized, so registration is assumed to correspond to maximizing the mutual information [31].

LBFGSB Optimizer

Limited memory Broyden-Fletcher-Goldfarb-Shannon minimization (LBFGS) is an algorithm for solving non-linear optimizations which is based on the gradient projection method [11]. Limited memory Broyden-Fletcher-Goldfarb-Shannon minimization with simple bounds extends LBFGS to allow specific bounds to be set for the parameters in the search space [4]. It is suited to applications which have a large number of parameters, which makes it a good choice for optimizing the parameters of the B-spline transformation at each iteration.

Deformation Vector Field

The deformable registration is most frequently used to transform images into the same coordinate space before making comparisons. However, the transform also outputs a dense vector field where the vector at each point in the field describes the difference between that point in the source image and the corresponding point in the target image. This provides a way to assess differences between the two images at every point. However, the information in the vector field can be difficult to interpret since many of the point displacements may be due to small or random differences. It is desirable to identify points where the differences between the images is likely to be interesting.

3.3.2   *Detecting Regions of Interest in the Deformable Vector Field*

To interpret the deformation vector field in a meaningful way, it is necessary to define which differences between two images are significant. To detect regions of significant differences, low-level and mid-level features are extracted from the deformable registration vector field which will be used to identify and describe regions in the difference image representing shape differences of interest.

Low-Level Vector Field Properties

Three vector properties were chosen and evaluated independently. The properties selected are:

1. deformation vector magnitude,

2. cosine distance between the deformation vector and the surface normal vector,

3. cosine distance between the deformation vector and a predefined reference vector.

Each of these properties can be extracted from the vector field as a scalar image on which pattern analysis techniques can be applied. In addition, these properties can be combined into vector images so pattern analysis can be performed jointly.

Mid-Level Vector Field Properties

Once the low-level features have been extracted from the vector field, they can be used to calculate mid-level features that quantify organization in the vector field on the neighborhood level.

Average Neighborhood Difference Measure

The local similarity measure for this application is defined as the average difference between a point and its neighbors within a radius r. When the local similarity is calculated for a vector field property, the value at each point represents the difference between that point and its neighborhood average. A low value of average neighborhood difference indicates that the voxels surrounding the center point are similar in value. Figure 3.6 shows a comparison of K-means based clustering and the average neighborhood difference measure. The neighborhood difference measure can be used to indicate which of the clusters found by K-means have low local differences. In figure 3.6(b), the lowest levels of average neighborhood difference are shown in red. It can be seen that the region centered on the midface has the lowest value for this measure, indicating that the corresponding clusters in figure 3.6(a) have a higher level of organization. This is expected since the image difference was due to an applied warping known to be constant in this region.

Local Entropy Measure

Entropy can be interpreted as a measure of disorder or unpredictability, so it is a natural choice for a metric when looking for organization in a data set. Given N observations $\{x_1, x_2, ..., x_N\}$, which occur with probabilities $\{p_1, p_2, ..., p_N\}$, the Shannon entropy is defined as

$$H = \sum_{i=1}^{N} p_i log \frac{1}{p_i}. \tag{3.5}$$

(a) K-means based clustering of angle between deformation vectors and a reference vector

(b) Average neighborhood difference of angle between deformation vectors and a reference vector
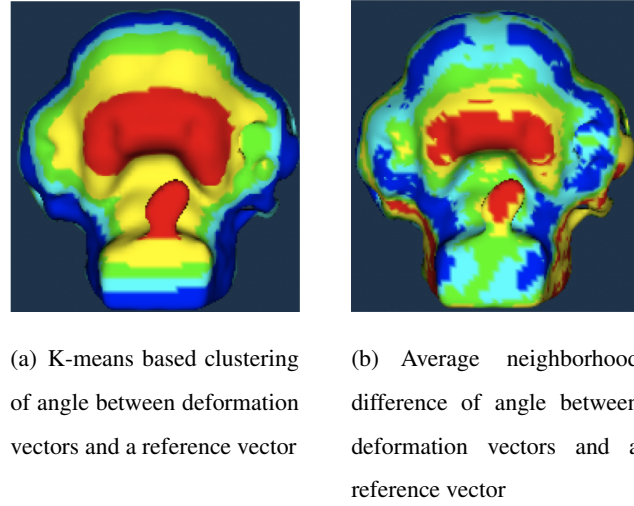
Figure 3.6: Comparing K-means based clustering of similar feature values to average neighborhood difference measure

Shannon entropy is inversely proportional to the log of the probability of observation i, which indicates that the less likely the observation is, the higher its entropy will be [34]. A set of observations that are equally likely will have maximal entropy. This distribution of observations corresponds to the highest level of unpredictability, since there are no dominant values. Conversely, a group of observations with higher predictability, that are clustered around a few values, will have lower entropy.

The local entropy measure is used at each point to calculate the entropy of a neighborhood with radius r centered on that point. Points with low entropy are centered on a neighborhood where the values contain a high degree of predictability or order. Points with higher entropy are located in neighborhoods where the value distribution is more dispersed, and therefore less ordered.

Chapter 4

VALIDATING THE VECTOR FIELD FEATURES ON SYNTHETIC DATA

To demonstrate the utility of the low-level and mid-level features introduced in the previous chapter, the features were evaluated on artificial data with predetermined differences. Each feature was independently clustered into regions of feature similarity and these regions were used to detect the known differences. The results of these experiments will be presented in this chapter.

4.1 Clustering and Merging Similar Regions

Once the deformable vector field features have been identified they are clustered to form regions using K-means clustering. For low-level features, these are regions of similarity. For mid-level features, the regions represent areas of consistent feature organization. Spatial constraints and cluster merging are then applied to these regions.

4.1.1 K-means Clustering of Similar Values

The K-means algorithm is a method for defining K classes that represent statistical distributions in a set of observations. For a group of N observations $\{x_1, x_2, ..., x_N\}$, K classes are found with means $m_1, m_2, ..., m_K$ such that

$$D = \sum_{i=1}^{N} \min_{1 \leq j \leq K} (x_i - m_i)^2 \tag{4.1}$$

is minimized. K-means was chosen for the task of clustering vector properties because it is computationally efficient and effective. One of the main challenges in using the conventional K-means algorithm is that the value of K needs to be estimated or known in advance. This problem was addressed by choosing a value of K significantly higher than the expected value and using a final cluster-merging step, described later in this section, to merge over-segmented regions. Figure 4.1 shows an example of clustering locations in a deformable vector field which have similar values of vector magnitude. Figure 4.1(a) shows a deformable vector field. In figure 4.1(b), the magnitude of

(a) Deformation vectors   (b) Deformation vector magnitude   (c) K-means clustering output
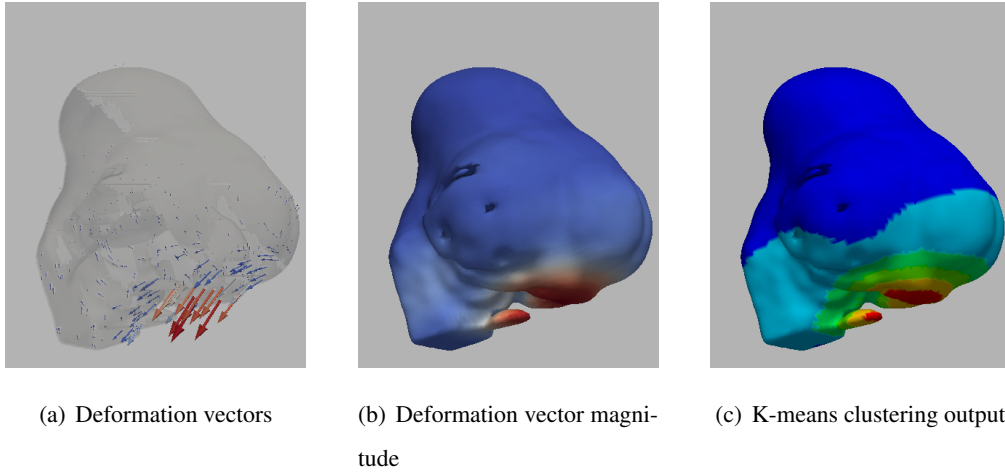
Figure 4.1: K-means clustering of deformation vector magnitude

these vectors are mapped to the surface of the target image, where the highest values are shown in red. Figure 4.1(c) shows the clusters these values have been assigned to, where each color represents a different cluster.

4.1.2   Spatial Constraints and Merging

Once K clusters of similar values have been segmented, a spatial constraint is imposed on the pixels. This is necessary since the goal of this method is to find spatially varying patterns. Voxels can be assigned to the same cluster if they have similar values even if they are spatially disconnected. Adding a spatial constraint allows clusters with disconnected regions to be split into their connected components. This process may result in new clusters with a very small number of voxels. Before proceeding to the linking step, a minimum cluster size threshold is applied and clusters with a size below this threshold are removed. This threshold is set with respect to the image resolution and the expected physical size of relevant shape change for a data set.

Once the spatially constrained clusters have been identified, an algorithm is applied to merge clusters that may have different means but are part of the same spatially varying pattern. The goal of this algorithm is to identify neighboring clusters where the voxels on the cluster borders have a high level of similarity. To accomplish this, the edge set of voxels from each cluster border is identified. Figure 4.2 shows an example of the edge set of a clustered image. Next, for each neighboring

(a) Deformation vectors magnitude clustering output

(b) Edge set of deformation vector magnitude clusters
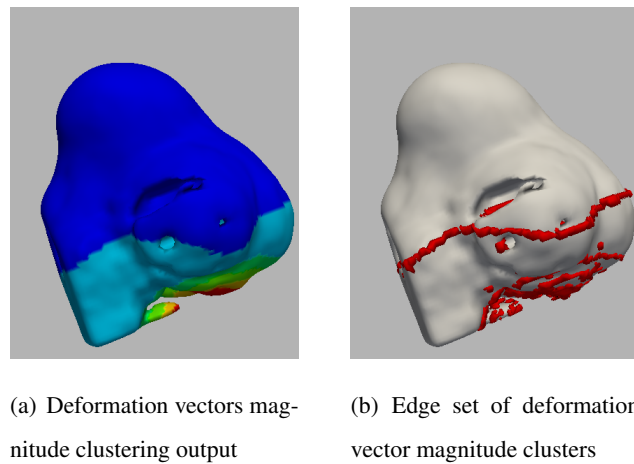
Figure 4.2: Edge set of a clustered image.

cluster, the subset of border voxels shared with this neighbor is found. The vector property values for neighboring voxels on each side of the border are compared to get a value expressing the similarity at the border of the two clusters. This is repeated for each cluster in the image. Finally, a similarity threshold is applied to the border similarity values computed for each set of neighboring clusters, and this determines whether they should be merged.

4.2 Describing Regions of Interest

Once the clusters have been identified, high-level cluster features describing the region are extracted. The following high-level cluster properties have been identified as useful for identifying and characterizing regions of significant difference:

- location of cluster center (x, y, z),

- number of voxels in cluster,

- entropy of cluster,

- average local entropy,

- average neighborhood difference,

- average magnitude of deformation vectors,

- average difference in angle between deformation vectors and surface normal,

- standard deviation of deformation vector magnitude and angle,

- standard deviation of local similarity.

These features can be used to decide if the region is significant and rank the regions in order of significance. They can also be used to describe the characteristics of detected patterns.

4.3 The Synthetic Data

Synthetic data was generated to test the effectiveness of the clustered feature regions in detecting relevant differences between images. One image of a chick embryo head was chosen as a reference and deformed to create images with predetermined differences. The synthetic images were deformed using landmark warping to create small magnitude deformations in regions critical to the development of the midface. In landmark warping, landmarks are identified in an image and their desired location in the deformed output image is specified. These transformed landmark points are then interpolated to define a transformation that is applied to the whole image.

Two sets of synthetic data were generated. The first consists of three images warped to closely mimic three typical varieties of midfacial defects. Figure 4.3 shows the initial landmark placement, interpolated warping magnitude, and deformation vectors for the three landmark-warped images created. These images were used to show the effectiveness of this method for specific deformations of interest.

A second synthetic data set was developed to provide a larger number of images with small magnitude deformations in a variety of locations across the midface. This was used to test the ability of the feature regions to spatially locate the disturbances introduced in the midface. The warped images were generated by placing a $5x5$ grid over the midface region. A warped image was generated for each grid square using 4 landmarks placed in that square. This resulted in a set of 25 images, each with a small magnitude deformation in a distinct and known location.

Frontal view of land-
mark placement and
warping magnitude
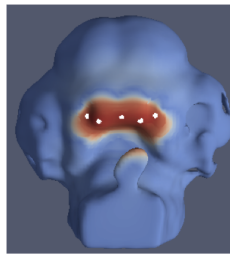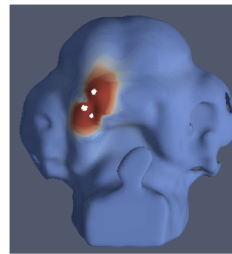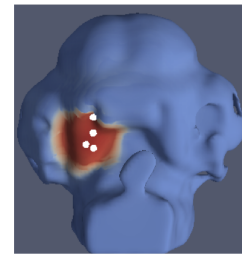
(a) Image 1   (b) Image 2   (c) Image 3

Lateral view of defor-
mation vectors
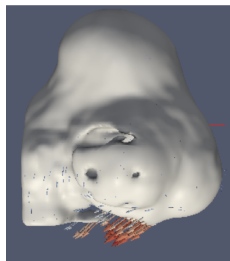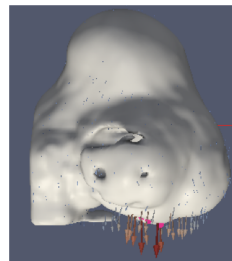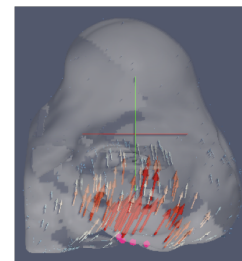
(d) Image 1   (e) Image 2   (f) Image 3

Figure 4.3: Figures 4.3(a) - 4.3(c) show landmark points marked in white. Image color indicates the interpolated transform magnitude, with highest values shown in red and lowest in blue. The deformation vectors for these three images are shown in figures 4.3(d) - 4.3(f).

4.4   Experimental Results

In this section, example results from the K-means clustering method of vector characteristics, average neighborhood similarity measure, and local entropy measure are presented for the synthetic data. These results validate the use of the low-level, mid-level, and high-level cluster features introduced in this work.

4.4.1   K-means Clustering of Low-Level Vector Characteristics

For these experiments, each landmark-warped image was registered to the original image, resulting in a vector field describing their differences. Vector characteristics were extracted from the fields and the K-means clustering method was applied with $K = 5$. The merging step was omitted so the cluster mean values could be compared across images and related to ground truth values. The features describing the differences between the warped images and the unwarped template image were evaluated to determine if they could accurately identify and describe the regions where these differences occurred.

First, the feature clusters were analyzed for the synthetic differences mimicking common mid-facial defects. Figure 4.4 shows examples of the results for clustering of the vector magnitude and the reference vector angle. In all figures, clusters with the highest mean value are shown in red and those with the lowest mean value are shown in blue. In figures 4.4(a) through 4.4(c) it can be seen that the vector magnitude clusters with the highest mean are located in areas of high magnitude of the original warping transforms, shown in figures 4.3(a) through 4.3(c). In figures 4.4(d) through 4.4(f), the reference vector angle values find clusters around similar regions. The area incorporated is larger, since it includes regions where the transform magnitude is lower but the angle is similar as it is interpolated across the image. In these images the cluster means represent direction rather than magnitude. The first two images show the cluster around the warped region in red. The third image shows this region in cyan, indicating the different direction of its landmark warping. Normal vector angle results are not shown, since they are not suited to landmark-warped images, which are deformed without respect to surface curvature.

To further validate these findings, the second synthetic data set was used to show the ability of this method to locate the region effected by the deformation. To accomplish this, the average cluster

Magnitude clusters
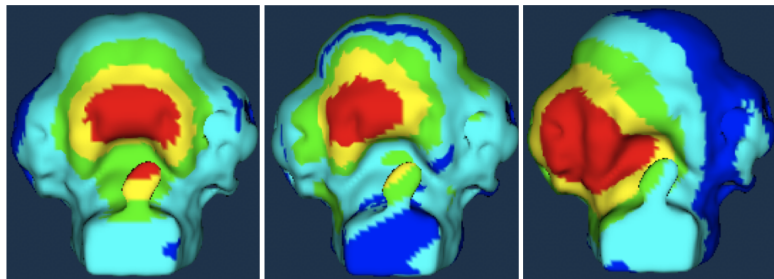
(a) Image 1 (b) Image 2 (c) Image 3

Reference vector angle
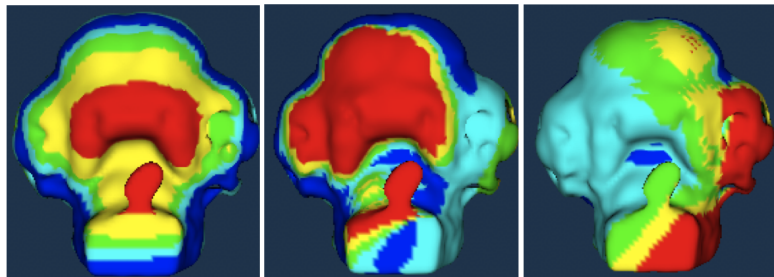
clusters

(d) Image 1 (e) Image 2 (f) Image 3

Figure 4.4: Frontal view of magnitude and reference vector clusters for the three landmark-warped images shown in figure 4.3.

Table 4.1: Relationship Between Vector Magnitude Clusters and Landmark Warping Transforms Used to Deform the Images

Cluster Rank	Mean Warping Transform Average	Mean Warping Transform STD	Landmarks in cluster (100 total)
1	2.026	0.577	100
2	0.474	0.217	0
3	0.206	0.069	0
4	0.087	0.025	0
5	0.022	0.004	0

magnitude, one of the high-level features, from the vector magnitude clusters was compared to the transformation used to create the data set. The cluster with the highest mean magnitude value is expected to occur where the magnitude of the landmark transformation used to deform the original image is the highest. Table 4.1 shows the ranked magnitude clusters, mean value of the landmark transformation magnitude for the cluster regions averaged across the 25 regions, and the standard deviation of the mean values. It also indicates the total number of deformation landmark locations present in each cluster. It can be seen that the cluster rank corresponds to increasing average landmark transformation magnitude. Furthermore, the cluster with the highest rank identifies the location of all the landmarks used to define the transformations.

4.4.2 Neighborhood Similarity Measure

Figures 4.5(a) through 4.5(c) show results from the neighborhood similarity measure for the first synthetic data set. The neighborhood similarity measure is shown applied to the reference vector values for each of the three warped images. High values of neighborhood similarity indicating regions of interest are shown in red. Clusters with low values are shown in blue. It can be seen that the landmark-warped regions have high levels of neighborhood similarity in each case, with the highest values located near the landmark points. This is expected since the warping transform applied in that region had a high level of consistency, resulting in high local organization.

Neighborhood similarity of reference vector angle clusters

(a) Image 1

(b) Image 2

(c) RImage 3

Local entropy of reference vector angle clusters

(d) Image 1

(e) Image 2

(f) Image 3
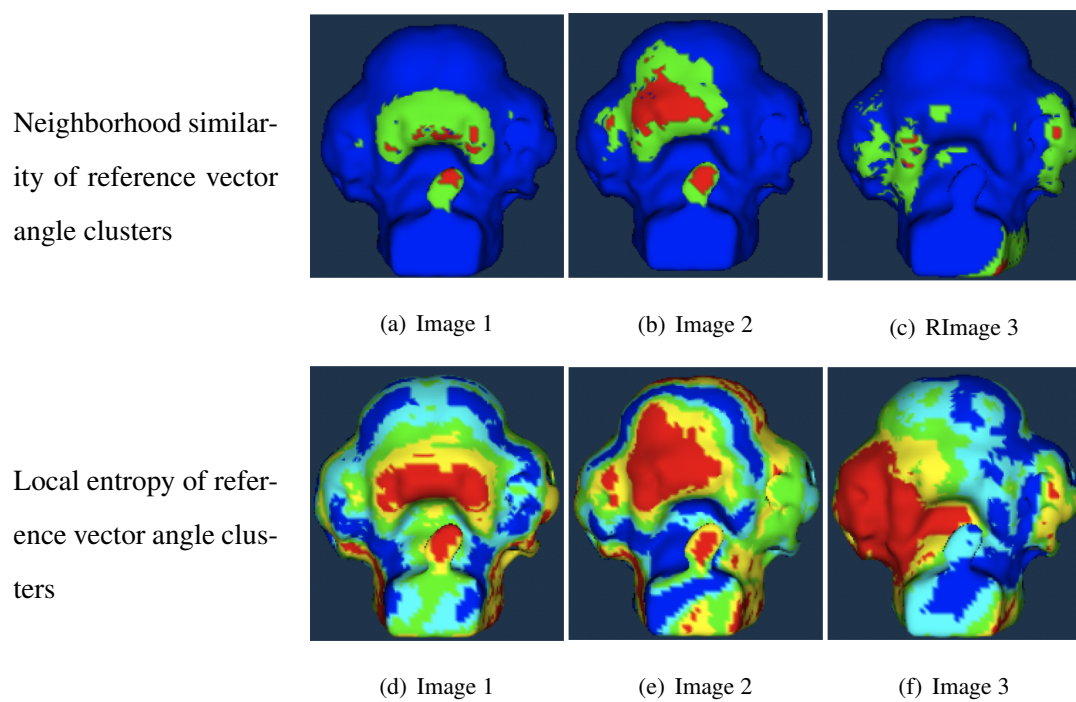
Figure 4.5: Frontal view of neighborhood similarity and local entropy of reference vector angles for the three landmark-warped images shown in figure 4.3.

Table 4.2: Summary of Feature Clustering Results on Landmark-Warped Images

Feature	Results
Magnitude	Landmark-warped regions identified by clusters with highest mean
Reference Vector Angle	Regions similar to magnitude identified, mean represents direction
Normal Vector Angle	Not suited to landmark warping
Local Similarity	Highest mean clusters located in landmark-warped region
Local Entropy	Landmark-warped region identified by clusters with lowest mean

4.4.3 Local Entropy Measure

Figures 4.5(d) through 4.5(f) show results from the local entropy measure applied to the reference vector angle values for first synthetic data set. For local entropy, low values indicate regions of interest, so clusters with the lowest mean values are highlighted in red and clusters with the highest mean values are shown in blue. It can be seen that the landmark-warped regions are located within the lowest entropy clusters in each case, correctly identifying the regions of high consistency resulting from the applied deformation.

4.4.4 Discussion of Results

The experimental results, summarized in table 4.2, show that the vector field features, K-means clustering method, average neighborhood difference measure, and local entropy measure are capable of detecting and describing regions of significant difference between two images. This validates the use of these features in quantifying and characterizing image distances.

Chapter 5

QUANTIFYING AND CHARACTERIZING ASYMMETRY IN BILATERALLY PAIRED STRUCTURES

Craniofacial microsomia is a condition resulting in marked asymmetry of the face. Research into the causes of craniofacial microsomia has been considerably hindered by the lack of appropriate animal models with the condition. However, researchers at SCRI have recently identified a number of new mutant mouse lines that exhibit fluctuating facial asymmetry involving both the upper and lower jaw and ears, similar to that seen in patients with craniofacial microsomia. These mice therefore represent excellent models for understanding the pathogenetic mechanisms of this condition. In mammals, the lower jaw bone is comprised of bilaterally paired hemi-mandibles. In humans, these fuse at the anterior midline by the end of the first year of postnatal life. However, in mice a cartilaginous join typically persists that allows the paired hemi-mandibles to be easily separated postmortem or digitally post-imaging.

The aim of this component of the work is to develop a tool by which asymmetry between hemi-mandibles and other bilaterally paired structures can be quantified and the spatial distribution of asymmetry assessed and compared across individuals. The tool introduced in this work will therefore aid research into the developmental basis of this condition and the specific role of the causative gene(s) and non-genetic factors.

To accomplish this aim, this chapter introduces a tool to quantify and characterize asymmetry in bilaterally paired structures. This method uses the deformable registration features and clustering method validated on synthetic data in the previous section to detect and describe regions of relevant asymmetry between two images of bilaterally paired structures. Three methods are provided to analyze the asymmetries. The *global asymmetry score* uses cluster features to quantify overall asymmetry. The *local asymmetry score* quantifies asymmetry in user-defined regions of interest. The *asymmetry similarity measure* quantifies pairwise similarity of individual asymmetry. The scores and image distances generated by this tool are shown to correlate highly with asymmetry

(a) Right Hemi-mandible   (b) Left Hemi-mandible
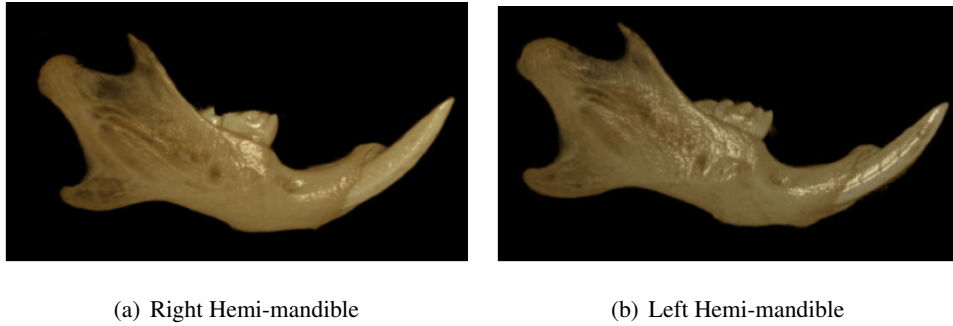
Figure 5.1: Rendered image of segmented left and right hemi-mandibles from the data set. The left hemi-mandible is mirrored to match right.

scores assigned by an expert through visual assessment of the image overlays.

5.1 Mouse Mandible Data Set

The data used to test the methods presented in this work consists of 23 microCT scan datasets of adult mouse mandibles ranging from subtly asymmetric to significantly asymmetric. An example of a segmented left and right hemi-mandible is shown in figure 5.1. The scans included a similar number of male and female animals, all between 28 and 30 days old to reduce effects of age on size and shape. The scans were performed at an isotropic resolution of 18 microns, then each dataset reduced by a factor of 3 to simplify data handling and computation time. To quantify the symmetry of the mandibles, the external contour of each hemi-mandible was first extracted from the image. Our method based on geodesic active contours was used to remove scan noise in the images, clarify indistinct borders between the object and the background, and to fill gaps and holes. Once the contours were extracted, the left hemi-mandibles were mirrored to permit comparison to the right sides and an affine transformation applied to align the images and remove pose differences.

5.2 Symmetry Analysis

The goal of the tool presented in this work is to provide a flexible way to analyze the asymmetries between bilaterally paired structures – in this case, the left and right hemi-mandible of individual mice – in a manner permitting ranking, local analysis, and comparison across individuals. This tool provides three primary modes of analysis:

1. the *global asymmetry measure* quantifies asymmetry across the surface of the mandible for an individual,

2. the *local asymmetry measure* quantifies asymmetry at user-defined regions of interest,

3. the *asymmetry similarity measure* quantifies similarity of asymmetry between different individuals.

5.2.1 Feature Generation and Description of Asymmetric Regions

As the first step in the symmetry analysis, the difference image describing the transformation between an specimen's right and mirrored left mandible is acquired. For this application, two low-level features were selected: 1) the deformation vector magnitude and 2) the cosine distance between the deformation vector and the surface normal vector. These two low-level features are clustered to find regions with similar transformation properties for each feature, using the K-means based method used in previous chapters. Asymmetric clusters are defined as those with an average deformation magnitude greater than T, where T is one standard deviation above the mean deformation magnitude of all voxels. The cluster features used to generate the feature vector are the:

1. location of cluster center,

2. number of voxels in cluster,

3. average magnitude value, and

4. average normal angle difference.

Each feature is normalized over the data set to remove differences in scale and distribution so that one feature does not dominate the feature vector.

5.2.2 Global Asymmetry Score

The first analysis method provided by this tool is the *global asymmetry score*. This score is used to quantify the overall asymmetry between the left and right hemi-mandibles, using features from the

deformation between them. The assessment of the feature properties would ideally correlate to the important asymmetry features considered visually by an expert, thus the automatic scores should correlate with the expert ranking of asymmetry. The expert ranking incorporates prior knowledge such as the relative importance of small regions of high magnitude differences and the need to exclude specific regions, such as the teeth because of variation due to wear. The global asymmetry score provides the user with the option to include this information in the automatic scores. The score is calculated using only deformation magnitude cluster features, since the direction of the deformation was not relevant to the ranking of asymmetry. The two global magnitude features used are:

1. U_{max} = maximum cluster mean,

2. V_{def} = total number of voxels in clusters with mean higher than the threshold T,

where T is one standard deviation above the mean deformation magnitude of all voxels. These global features are normalized over the data set and the score is defined as:

$$S_{global} = \alpha U_{max} + (1 - \alpha)V_{def}, \tag{5.1}$$

where α is a constant specifying the contribution of the maximum deformation relative to the overall deformation. The value of this constant is specific to the application and allows the user to incorporate prior knowledge about the relative importance of small regions with high magnitude transformations compared to larger regions with smaller magnitude transformations. In addition, the tool allows clusters centered in regions selected by the user to be excluded, for example, if deemed not to be biologically relevant to the particular question being considered.

5.2.3 Local Asymmetry Score

The *local asymmetry score* is a second asymmetry measure that provides an additional way to add prior knowledge to the asymmetry score. The local asymmetry score uses only features from the deformation vectors in a neighborhood around the landmark points placed by an expert. The local asymmetry score is defined as:

$$S_{local} = \alpha N_{max} + (1 - \alpha)N_{average}, \tag{5.2}$$

where N_{max} is the maximum neighborhood average, $N_{average}$ is the average deformation magnitude over all neighborhoods, and α is a constant specifying the significance of the maximum deformation versus overall deformation.

5.2.4 Similarity of Asymmetry

The third analysis method provided by this tool is a similarity measure to compare asymmetry across individuals, allowing individuals with similar shape and magnitude of asymmetry to be identified. This metric can be used on the difference clusters from a combination of feature types or the feature types can be evaluated independently. The similarity is calculated by comparing the set of cluster feature vectors for each mandible. To allow for maximum flexibility in query types, the user is also allowed to choose features from the feature vector to include for a specific query. To compare a mandible i to a second mandible j, the minimum distance from each cluster k in mandible i to a cluster in mandible j is found using:

$$c_{ik} = \min_{1 \leq n \leq n_j} d(f_{ik}, f_{jn}), \tag{5.3}$$

where $f_{i,k}$ is the feature vector for cluster k from mandible i, $f_{j,n}$ is the feature vector for cluster n from mandible j and n_j is the total number of clusters of mandible j. The total quality of the match between mandibles i and j is given by:

$$m_{i,j} = \frac{1}{n_i} \sum_k c_{ik} * magnitude_k, \tag{5.4}$$

where $magnitude_k$ is the element of the feature vector f_{ik} that describes the size of the cluster. The cluster matches are weighted by the size of the cluster in i to determine their relative importance to the overall image match. The final distance measurement between mandibles i and j is calculated by repeating this process to find $m_{j,i}$, the distance between mandibles j and i, and averaging the value of the matches from each direction,

$$Dist_{ij} = \frac{m_{i,j} + m_{j,i}}{2}. \tag{5.5}$$
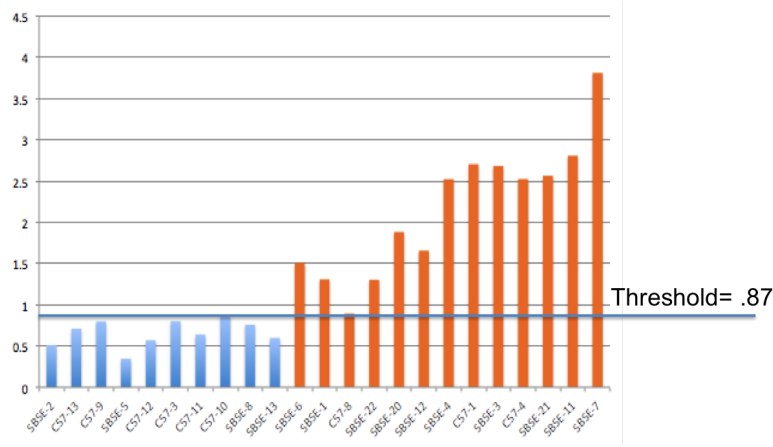
Figure 5.2: Global asymmetry scores for ranked mandibles. Blue indicates "minimal asymmetry" and orange indicates "significant asymmetry". Expert ranking is shown by the order of the X-axis.

5.3 Comparing Asymmetry Scores to Expert Rankings

In this section results are presented that motivate the use of the global and local asymmetry scores and the asymmetry similarity measure. When generating the global and local asymmetry scores the constant α was set to 0.75 and the region around the incisor tooth was excluded.

5.3.1 Global Asymmetry Score Evaluation

To provide ground truth for the global asymmetry score, an expert divided the mandible pairs into two groups, "minimal asymmetry" and "significant asymmetry", as indicated by the color of the bars in figure 5.2, then visually ranked each of the 23 mandible pairs in the dataset from most symmetric to least symmetric. This ranking is shown by the order on the X-axis in figure 5.2. We compared the global asymmetry scores of the two groups, indicated by the height of the bars in figure 5.2, and found that they could be separated by a threshold of 0.87. Furthermore, the correlation coefficient between the expert ranking and the global asymmetry scores was 0.92.

5.3.2 Local Asymmetry Score Evaluation

The local asymmetry scores were generated using features from connected points around four landmarks identified by an expert. These scores were evaluated using the same method as the global
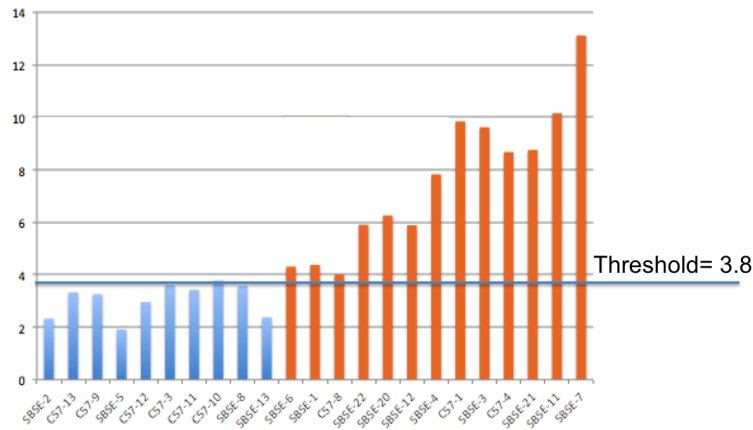
Figure 5.3: Local asymmetry scores for ranked mandibles. Blue indicates "minimal asymmetry" and orange indicates "significant asymmetry". Expert ranking is shown by the order of the X-axis.

asymmetry scores. The "minimal asymmetry" and the "significant asymmetry" groups could be separated using a local symmetry threshold of 3.8. The correlation coefficient between the expert ranking and the local asymmetry scores was 0.91 The local asymmetry scores are shown in figure 5.3.

5.3.3 Comparison to Landmark-based Method

The goal of this method is to quantify and describe asymmetry without the use of landmarks, but for completeness the global and local scores were compared to a landmark-based method, since an acceptable number of landmarks can be identified for this data set. For each hemi-mandible, 15 landmarks were placed by an expert. The left mandible from each pair was mirrored to match the right, and the hemi-mandible pairs of each individual were superimposed. The Procrustes distance, defined as the square root of the sum of squared distances between homologous landmarks, was used to rank mandibles in order of asymmetry. The correlation coefficient between this ranking and the expert ranking was 0.28, which indicates this method is not suitable for this application.

5.3.4   Similarity of Asymmetry Evaluation

The asymmetry similarity measure provides the distance from each mandible pair in the database to all other mandible pairs, with respect to either deformation magnitude or normal angle difference. Because the similarity measurement uses the spatial distribution of the asymmetry and because of the difficulty in incorporating spatial distribution in a simple expert ranking scale, they are not expected to correspond to the expert rankings in all cases. However, in the case of the magnitude asymmetry distances for the most asymmetric pair, it would be expected that these would better correspond to the difference in rank. In fact, the correlation coefficient between the expert rankings and distance from the most asymmetric pair is 0.91.

To motivate the usefulness of the magnitude and normal angle asymmetry comparisons, two sample queries are shown. In figure 5.4 a sample query is shown for the magnitude clusters, using the cluster center and average magnitude features. The query image chosen is the most asymmetric case: 5.4(a). In the left/right overlay in figure 5.4(d), the left hemi-mandible, shown in blue, is longer than the right hemi-mandible, shown in red. The top two matches for the magnitude of asymmetry are also shown in figures 5.4(b) and 5.4(c). Both show a similar magnitude of asymmetry in the condylar and angular processes. These regions are also circled in figures 5.4(d), 5.4(e), and 5.4(f). Note that in the result images, the left mandible is shorter than the right. This is expected since the direction of the asymmetry is ignored and only the magnitude of the flow vectors is considered. This type of query can be used to find mandibles with asymmetries similar in magnitude but independent of left/right directionality.

In figure 5.5, the sample query is shown for the normal angle clusters, using the cluster center and average normal angle difference features. The query image is the same as in the previous example and the top two matches for normal angle asymmetry are shown, but notably are different to those identified by the magnitude query. Compared to the magnitude query results, the difference is much smaller for the result image. However, in both normal angle results, the left hemi-mandible is shorter than the right, as in the query hemi-mandible pair. This type of query can be used to find asymmetries that have similar directionality despite differences in magnitude.

(a) Query   (b) First result   (c) Second result

(d) Query   (e) First result   (f) Second result
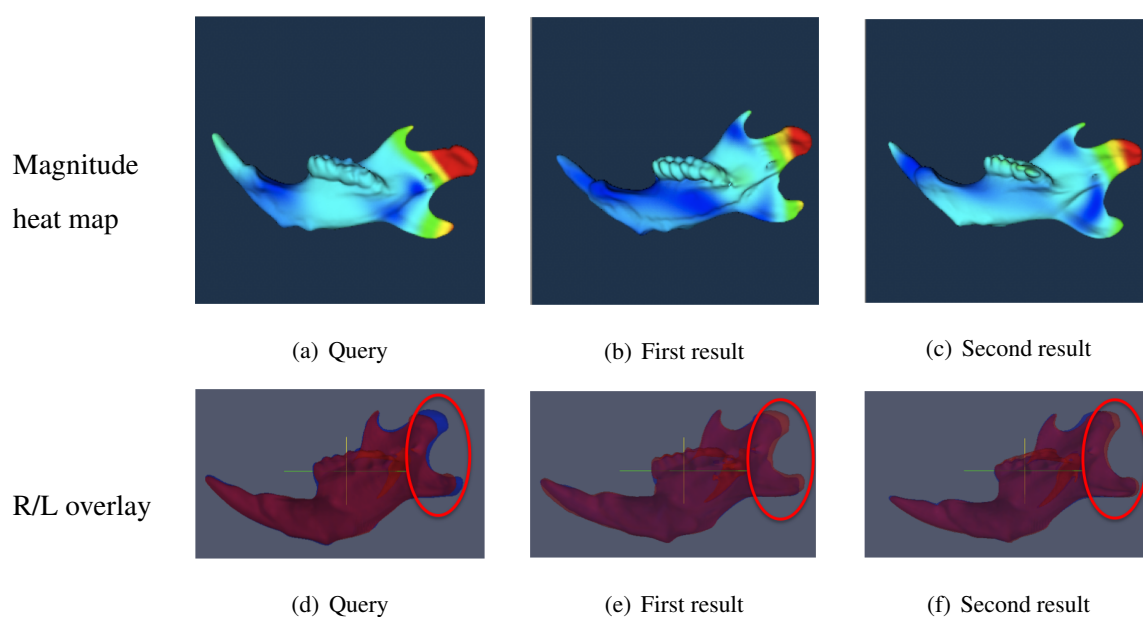
Figure 5.4: Heat maps and left/right overlay from a sample magnitude query using the most asymmetric mandible. In the top row the high magnitude is represented by red and the low by blue. In the lower row the right hemi-mandible is shown in blue, the left in red, and regions of asymmetry in the condylar and angular processes are circled.

Normal angle heat map

(a) Query   (b) First result   (c) Second result

R/L overlay

(d) Query   (e) First result   (f) Second result
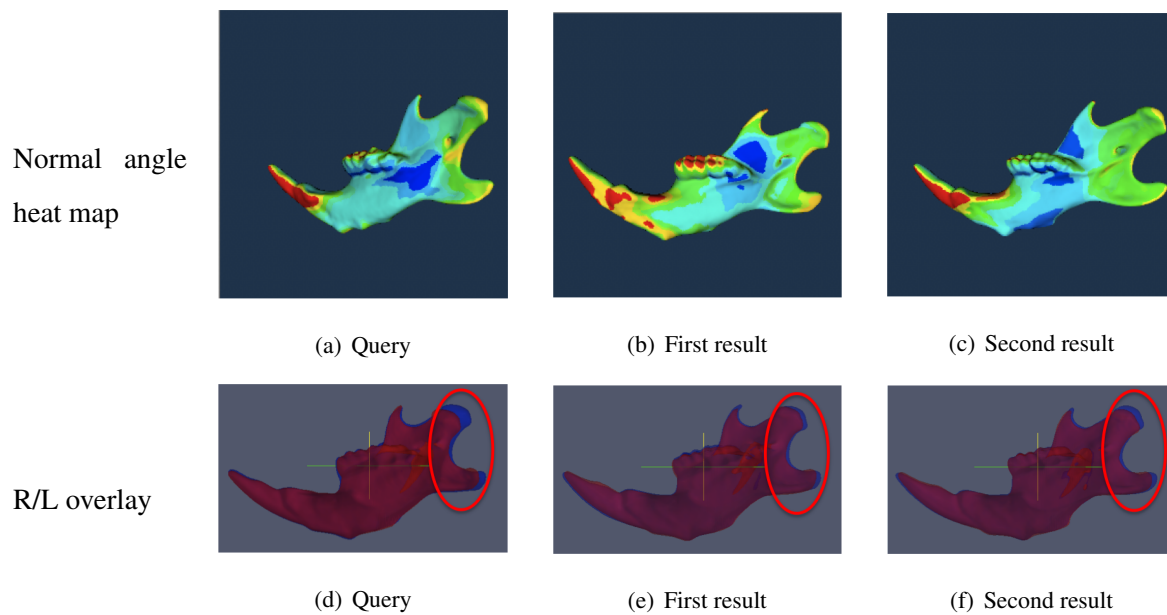
Figure 5.5: Heat maps and left/right overlay from a normal angle difference query using the most asymmetric mandible. In the top row the largest normal angle difference is represented by red and the lowest by blue. In the lower row the right hemi-mandible is shown in blue, the left in red, and regions of asymmetry in the condylar and angular processes are circled

Table 5.1: Summary of Asymmetry Scoring Results

Method	Correlation to Expert Ranking
Global Asymmetry Score	0.92
Local Asymmetry Score	0.91
Similarity to most asymmetric case	0.91
Procrustes-based method	0.28

5.4 Discussion

The new tool introduced in this chapter is capable of assigning asymmetry scores based on global features and features from user-defined locations. Regions of significant asymmetry are detected, described, and used to quantify the similarity of asymmetry across individuals. The evaluation of these methods on mouse mandibles with varying amounts of asymmetry is summarized and compared to the Procrustes-based method using landmarks in table 5.1.

An online user interface was developed to allow this tool to be accessed by multiple researchers. The interface allows the modules of the tool to be flexibly combined to support a wide range of applications and workflows. Figure 5.6 shows the interface set up to quantify the asymmetry in a pair of mouse hemi-mandibles.
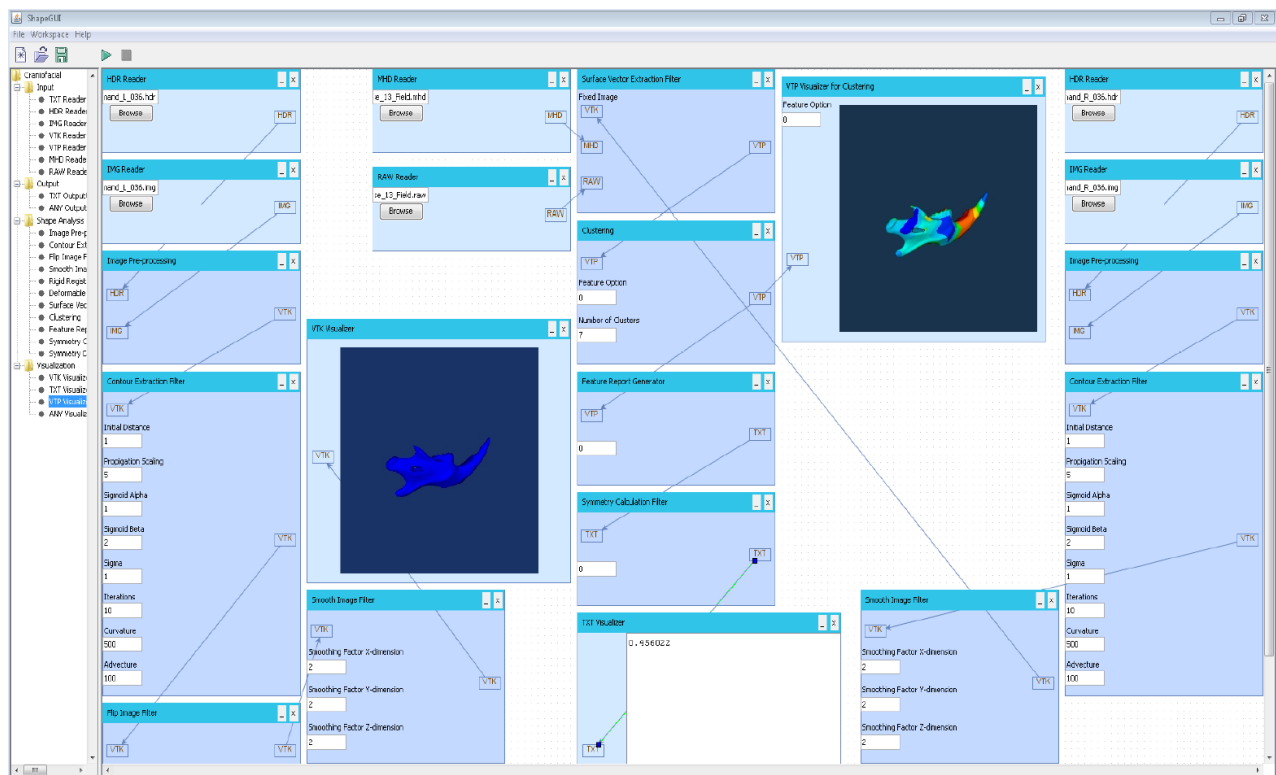
Figure 5.6: Online interface for asymmetry analysis. The task shown in this figure is the quantification of asymmetry in a pair of mouse hemi-mandibles.

Chapter 6

A SPATIOGRAM-BASED FRAMEWORK TO REPRESENT IMAGE DIFFERENCES

This chapter describes the framework developed to represent and compare the deformable vector field features that were experimentally validated in the previous chapter. This framework utilizes a spatiogram-based representation of the regions of difference to allow for the simultaneous comparison of feature values and their spatial distributions across difference images. A distance measure is introduced to quantify these comparisons that is capable of using single or multiple feature types.

6.1   Spatiograms

Histograms occupy a fundamental place in image analysis. They provide a robust description of image data which has proven to be powerful in detecting similarity between images. However, the loss of all spatial information in this representation poses a significant drawback in applications where the spatial distribution of values in an image is important. To address this deficiency, Birchfield and Rangarajan proposed a modified histogram that includes spatial information, referred to as a spatiogram [2]. For an image I, the histogram of I was defined as:

$$h_I(b) = n_b, b = 1, ..., B,$$ (6.1)

where n_b is the number of pixels with values assigned to the b-th bin and B is the total number of bins. The spatiogram of I was defined as the second order histogram:

$$h_I(b) = < n_b, \mu_b, \Sigma_b >, b = 1, ..., B,$$ (6.2)

where μ_b and Σ_b are the mean vector and covariance matrices of the spatial coordinates assigned to bin b. This modification allowed the spatiogram to preserve information about the spatial distribution of the points assigned to each bin by modeling it as a Gaussian distribution. However, this representation may be overly simplified for images where the bins may contain multimodal spatial

distributions. The original introduction of spatiograms suggested this could be addressed in the future by modeling spatial information with a mixture of Gaussians for each bin [2]. This approach was taken in [19] where Lyons defined the Mixture of Gaussian (MoG) spatiogram as:

$$h_I(b) = < n_b, m_b = ((\alpha_{b1}, \mu_{b1}, \Sigma_{b1}), ..., (\alpha_{bm}, \mu_{bm}, \sigma_{bm})) >, \tag{6.3}$$

where α_{bi} is the weight of the ith mixture and μ_{bi} and Σ_{bi} are the ith mixture parameters. One drawback of the variant proposed in this paper is that it required a constant number of distributions in each bin for the calculation of the distance between two spatiograms. In this work, we introduce a formulation of the MoG spatiogram where the number of distributions in each bin is allowed to vary.

6.2 Comparing Spatiograms

In Birchfield and Rangarajan's original paper on spatiograms, a distance metric between two spatiograms was proposed based on the Mahalanobis distance [2]. An improvement to this metric was proposed in [8] where a normalized distance derived from the Bhattacharyya coefficient was introduced. This metric exploited the fact that since a spatiogram is a 2nd order histogram, it could be converted back to a histogram with an extra dimension of space. Each bin b had its contents n_b divided over an infinite number of spatial bins $n_{b,k}$ where $[k] = [-\infty, +\infty]$. This was expressed as:

$$n_{b,k} = \frac{n_b \phi_b(k\Delta w)\Delta w}{\sum_{i=-\infty}^{+\infty} \phi_b(i\Delta w)\Delta w}, \tag{6.4}$$

where Δw is the spatial size of each bin and ϕ_b is a normalized Gaussian with the mean and covariance of bin b. Using this conversion, spatiograms could then be compared using the Bhattacharyya coefficient. The Bhattacharyya coefficent measures the amount of overlap between two statistical samples, defined for two discrete histograms n and n' as

$$\rho(n, n') = \sum_{b \epsilon B} \sqrt{n(b)n'(b)}. \tag{6.5}$$

In the case where n and n' are two spatiograms expressed as histograms with an added dimension this becomes:

$$\rho(n, n') = \sum_{b=1}^{B} \sum_{k=-\infty}^{+\infty} \sqrt{n_{b,k}n'_{b,k}} \tag{6.6}$$

$$= \sum_{b=1}^{B} \sum_{k=-\infty}^{+\infty} \sqrt{\frac{n_b \phi_b(k\Delta w)\Delta w}{\sum_{i=-\infty}^{+\infty} \phi_b(i\Delta w)\Delta w}} \sqrt{\frac{n'_b \phi'_b(k\Delta w)\Delta w}{\sum_{i=-\infty}^{+\infty} \phi'_b(i\Delta w)\Delta w}}. \tag{6.7}$$

This equation was simplified in [8] and the difference between two spatiograms h and h' was expressed as the spatially weighted sum of similarities:

$$\rho(h, h') = \sum_{b=1}^{|B|} \Psi_b \sqrt{n_b n'_b}, \tag{6.8}$$

where the normalized probabilistic spatial weighting term was defined as:

$$\Psi_b = 2(2\pi)^{0.5} |\Sigma_b \Sigma'_b|^{0.25} N(\mu_b; \mu'_b, 2(\Sigma_b + \Sigma'_b)) \tag{6.9}$$

and $N(x; \mu, \Sigma)$ denotes a Gaussian with mean μ and variance Σ evaluated at x. This normalized Bhattacharyya metric was modified for the MoG spatiogram distance in [19] to accommodate the m_b component. The distance between two MoG spatiograms was defined as:

$$\rho^{mm}(h, h') = \sum_{b=1}^{|B|} \Psi_b^{mm} \sqrt{n_b n'_b}, \tag{6.10}$$

where the normalized spatial weighting term for bins with m mixtures was:

$$\Psi_b^{mm} = \sum_{i=1}^{m} \alpha_{bi} \sum_{j=1}^{m} \alpha_{bj} \eta_{bij} N(\mu_{bj}; \mu'_{bi}, 2(\Sigma_{bi} + \Sigma'_{bj})) \tag{6.11}$$

and

$$\eta_{bij} = 2(2\pi)^{0.5} |\Sigma_{bj} \Sigma'_{bi}|^{0.25}. \tag{6.12}$$

In this formulation, m was required to be fixed for the bins being compared. Spatiograms have been primarily used for tracking color patches in video images, and for these applications a single or fixed number of Gaussians has been sufficient [19]. Since the goal of the framework introduced in this chapter is to provide an accurate description of complex shape changes over a larger region, these approaches could overgeneralize the feature descriptions. Ideally, the method for assigning the number of Gaussians to a bin should utilize information about the spatial distribution of the contents of that bin. To accomplish this, a connected components algorithm is used to estimate the number of spatially separate distributions in a bin and initialize the number of Gaussians before parameter estimation. A new variable MoG spatiogram metric is now introduced that accommodates

this flexibility in the number of Gaussians per bin. The one-directional difference between two MoG spatiograms h and h' is defined as:

$$d^{mn}(h, h') = \sum_{b=1}^{|B|} \Psi_b^{mn} \sqrt{n_b n_b'},$$
(6.13)

where

$$\Psi_b^{mn} = \sum_{i=1}^{m} \alpha_{bi} \max_{j \in [n]} (\eta_{bij} N(\mu_{bj}; \mu_{bi}', 2(\Sigma_{bi} + \Sigma_{bj}'))),$$
(6.14)

m and n are the respective numbers of distributions, $[n] = [1, ..., n]$, and

$$\eta_{bij} = 2(2\pi)^{0.5} |\Sigma_{bi} \Sigma_{bj}'|^{0.25}.$$
(6.15)

To achieve a symmetric distance measure, the total distance between h and h' is defined as:

$$\rho^{mn}(h, h') = \frac{(d^{mn}(h, h') + d^{mn}(h', h))}{2}.$$
(6.16)

This distance metric is symmetric, is normalized over the mixture components, and has a maximum value of $\rho(h, h') = 1$ when $h' = h$.

6.3   MoG Spatiogram-based Framework

One goal of this work is to provide a framework that can provide both quantitative descriptions between two images and compare the flow patterns between two sets of images. For example, in an application where the differences between images are due to developmental growth, the framework should provide both a method to describe the differences between images at two developmental stages and a method to quantify the differences in growth trajectories, such as a comparison of normal growth and growth impacted by a genetic abnormality.

6.3.1   Representing Feature Regions using MoG Spatiograms

The use of MoG spatiograms to represent the regions of feature differences supports comparison of flow patterns between two images and provides an effective and intuitive method to define distances that is capable of combining of feature types. In this framework, a MoG spatiogram will be used to represent the regions of difference for each feature type. High-level features of the MoG spatiogram will quantify characteristics of the image difference and the MoG spatiogram metric introduced in the previous section will define the distance between the flow patterns of two sets of images.

| (a) | (b) | (c) |
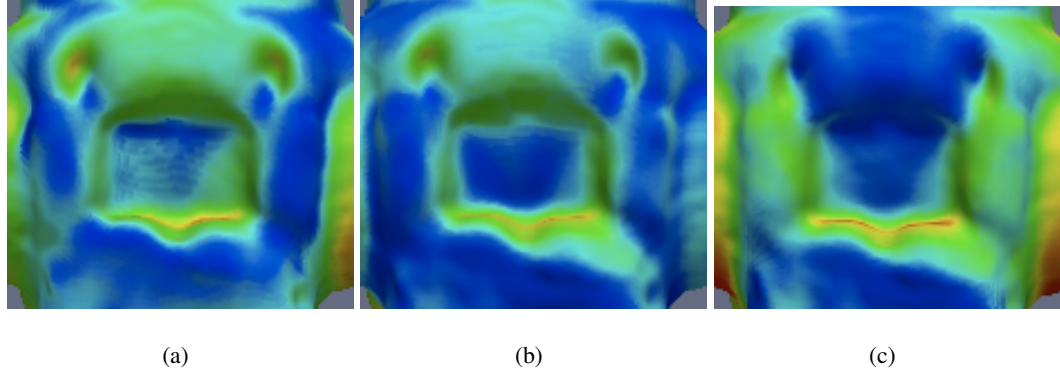
Figure 6.1: Normal angle heat maps from three difference images with similar histograms. Spatial information is needed to differentiate spatial distribution of values.

Constant Value Partitions with Spatial Information

In the previous chapter, the deformable vector field features were validated by using a K-means based clustering method to find regions of significant difference between two images. This clustering method provided regional partitions optimized for the values in each individual image for identification of significant regions of difference. One primary reason for using MoG spatiograms to represent the feature regions is that they provide histogram-based partitions that remain constant across images. This produces regions that are less dependent on the value distribution within an individual image and improves the comparison of regions across images. While spatiograms provide some of the benefits of image histograms, they also avoid the loss of spatial information about the location of the feature regions. The motivation for including spatial information can be seen in the example in figure 6.1. Three normal angle heat maps from vector flow images with similar histograms are shown. The addition of spatial information is necessary to distinguish the similar value distributions in figures 6.1(a) and 6.1(b) from the distribution in figure 6.1(c).

Integrating Feature Types

The MoG spatiogram distance metric introduced in this work allows for efficient combination of multiple feature types. Each feature type has a different value range which needs to be normalized before they can be combined. This is accomplished by first computing the MoG spatiogram distance

for each feature type of interest. These distances are combined using the weighted mean:

$$\hat{\rho}^{mn}(h, h') = \sum_{i=1}^{f} w_i \rho_i^{mn}(h, h'), \tag{6.17}$$

where

$$\sum_{i=1}^{f} w_i = 1, \tag{6.18}$$

f is the number of feature types, and $\rho_i^{mn}(h, h')$ is the MoG spatiogram distance for feature type i. The weights allows for the optional inclusion of a priori information about the relative importance of each feature to the specific application. In the absence of this prior knowledge, w_i is set to $1/f$ and the weighted mean reduces to:

$$\hat{\rho}^{mn}(h, h') = \frac{1}{f} \sum_{i=1}^{f} \rho_i^{mn}(h, h'). \tag{6.19}$$

Chapter 7

NORMAL CHICK EMBRYO DEVELOPMENTAL SERIES

In this chapter, the spatiogram-based framework introduced in the previous chapter is applied to assess pairwise morphological shape differences in normal chicken embryo facial development. The goal of these tools is to provide quantitative information about shape differences which are not currently known. To motivate the use of these tools to accomplish this goal, they will first be applied to differences due only to the embryonic developmental stage, which are known. A method for creating and displaying an average growth trajectory model is introduced to summarize the differences between two developmental stages and is used to further validate the use of these tools to find similarity between embryonic growth trajectories.

7.1 Data Set

The data set used for these experiments is a set of 16 optical projection tomography (OPT) images of normal chick embryos from five developmental stages. An example image of a developing chick embryo and a zoom of the facial region is shown in figure 7.2. Since only the midface was of interest in these experiments, the midface was defined using the alignment landmarks and this region was cropped from the chick embryo images.

This is a complex data set due to the variation within developmental stages and the variable amount of change between stages. A representative image from each developmental stage is shown in figure 7.2. It can be seen that the differences between some stages is large, while the difference between others is relatively subtle. Images can have shape or size similarities to one group, but be classified to a second group due to small developmental differences.

7.2 Retrieval of Similar Growth Trajectories using Spatiogram Distances

With the MoG spatiogram representation of shape difference features and the new spatiogram distance measure introduced in the previous chapter, it is now possible to express the similarity between

(a) Chick embryo with facial region highlighted

(b) Zoom of facial region
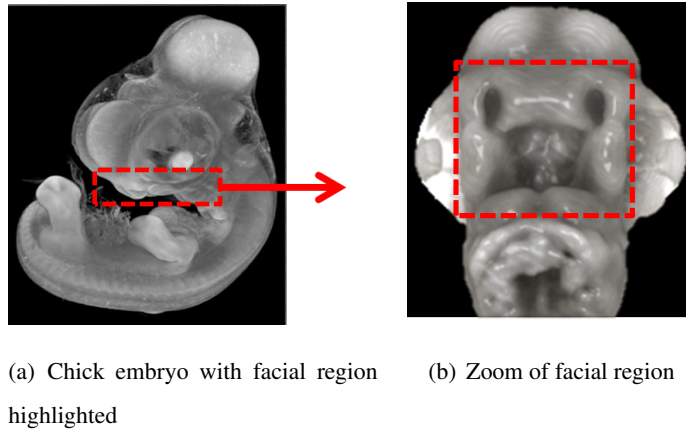
Figure 7.1: Developing chick embryo with facial region highlighted.

(a) HH 19.5　　(b) HH 24　　(c) HH 24.5　　(d) HH 25　　(e) HH 26
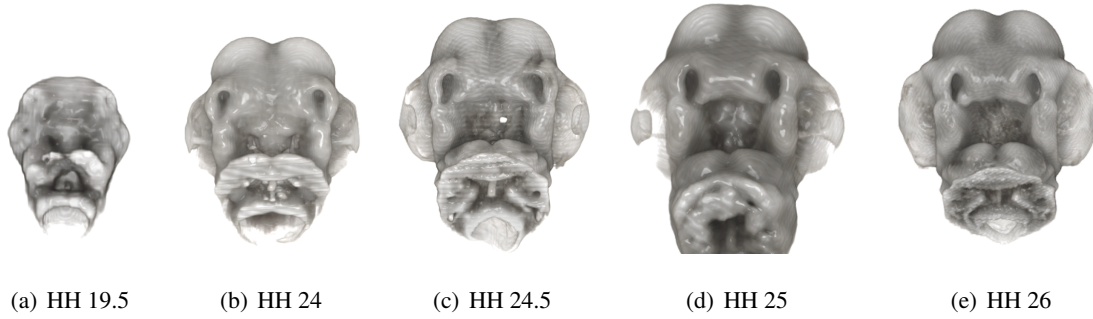
Figure 7.2: Examples of midface development for each developmental (Hamburger and Hamilton [14]; HH) stage in the dataset.

two difference images using a single distance value. These distance values can be used to perform similarity-based retrieval of difference images. Applied to the data set of normal chick embryos, the query difference image will represent a trajectory of growth between two developmental stages. Similarity between growth trajectories is evaluated using the spatiogram distance measure. Results are retrieved from the database by ranking them in order of the spatiogram distance from the query. It is expected that the most similar trajectories of growth should be between the same two developmental stages as the query difference image. To accomplish this, a representative template image was selected for a developmental stage and the MoG spatiograms describing the pairwise difference between the template image and each embryo in the data set were obtained. For this experiment

the normal angle and magnitude features were used for retrieval with no prior knowledge incorporated, so the spatiogram distances for each feature type were weighted evenly. Each of the resulting difference images was used as a separate query (omitting stage 21 due to the small sample size). The difference images in the dataset were ranked by the spatiogram distance from the query. An example of the feature heat maps for one such query and the top ranking results is shown in figure 7.3. In this figure and going forward, the normal angle and magnitude heat maps are differentiated by different color mappings for clarity.

(a) Query   (b) First result   (c) Second result   (d) Third result

(e) Query   (f) First result   (g) Second result   (h) Third result
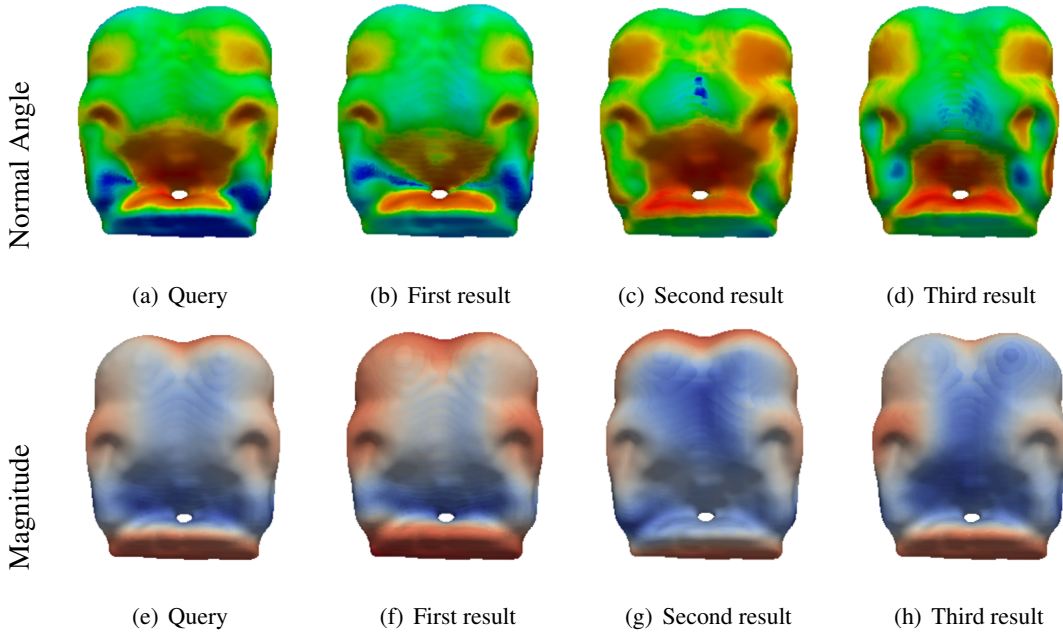
Figure 7.3: Normal angle and magnitude heat maps of MoG spatiograms for a sample query and the 3 top ranked results. The sample query and all 3 results represent growth trajectories between HH 24 and HH 26. Only the midface region is shown.

The success of each query was scored using the average normalized rank of the relevant difference images [23]. In this case, the relevant difference images are those representing a trajectory of growth between the same two stages as the query. The evaluation score for a query q is defined as:

$$score(q) = \frac{1}{N \cdot N_{rel}} \left(\sum_{i=1}^{N_{rel}} R_i - \frac{N_{rel}(N_{rel} + 1)}{2} \right),$$

(7.1)

Table 7.1: Average Retrieval Scores for Differences Between Stages

Templates	HH 24 Average Score	HH 24.5 Average Score	HH 25 Average Score	HH 26 Average Score
HH 19.5	0.087	0.018	0.156	0.020
HH 24	X	0.017	0.021	0.045
HH 24.5	0.044	X	0.008	0.069
HH 25	0.007	0.100	X	0.072
HH 26	0.030	0.067	0.045	X

where N is the number of objects in the database, N_{rel} is the number of database objects that are relevant to the query object q, and R_i is the rank assigned to the ith relevant object. The evaluation score ranges from 0 to 1, where 0 is the best score, indicating that all relevant database objects were retrieved before all other objects in the database.

The query scores for each developmental stage were averaged and this was repeated for each developmental stage template. The average evaluation scores by developmental stage are shown in table 8.1, where each entry in the table represents the average query score for a growth trajectory between two stages. The overall average score was 0.049, which is very close to the ideal score of 0, validating the methods used. Some incorrectly ranked retrievals are expected due to the variation within developmental stages.

All the scores in table 8.1 are very low with the exception of the score for the transitions between the HH 19.5 template and HH 25 embryos, which is slightly higher. This is a good example of where prior knowledge could be incorporated in the distance measure. Since these stages are far apart and are expected to have large size differences, the magnitude features are known to be highly significant. Using the magnitude features alone, the score for this category drops from 0.156 to 0. Increasing the weight of the magnitude features in cases where there are known to be large size differences can improve performance.

In addition to validating the methods in this section, this type of a retrieval is useful for classification purposes. In the case of an embryo image where the developmental stage is unclear or

unknown, the image can be warped to standard template images from known developmental stages. These growth trajectories can then be used as queries to find the most similar growth trajectories to the standard template images, identifying the most likely developmental stage of the unlabeled embryo image. Agreement across the standard template retrievals will reinforce the proposed label.

7.3 Average Growth Trajectories

To summarize the differences between developmental stages, an average growth trajectory was defined as the average transformation of all embryos from one developmental stage to a representative template image from a second stage. Representative templates were chosen over the use of average images to eliminate concerns about the biological relevance of average images in this application where the variation within groups is not well documented. These average growth trajectories are helpful in providing a visual representation of the average feature values which can be quickly and easily interpreted by researchers. In addition, statistics about the group, such as the standard deviation of the features, can be plotted on the image surface so any within-group variation can be visualized. An example of the average growth trajectory feature maps and their standard deviation maps is shown in figure 7.4. It shows the average of the growth trajectories between HH 24 and HH 26 and can be compared to individual growth trajectories between these two stages shown in figure 7.3. In addition to providing a useful visual summary, these average growth trajectories can also be used for retrieval.

7.4 Retrieval of Relevant Average Growth Trajectories

The second retrieval experiment demonstrates that individual growth trajectories between two stages are more similar to the average growth trajectory between the same two stages than to any other average growth trajectory. To show this, an average growth trajectory was generated for every possible transition between two developmental stages in the data set. To avoid having the query image contained in any of the average growth trajectories in the dataset for that query, hold-out averages excluding the query were generated for the relevant averages. Due to the small sample size of some developmental stage groups and the need for hold-out averages, the queries were chosen to be difference images representing the trajectories between stage HH 24 and all other stages. Stage 24

(a) Normal angle   (b) Magnitude

(c) Normal angle   (d) Magnitude
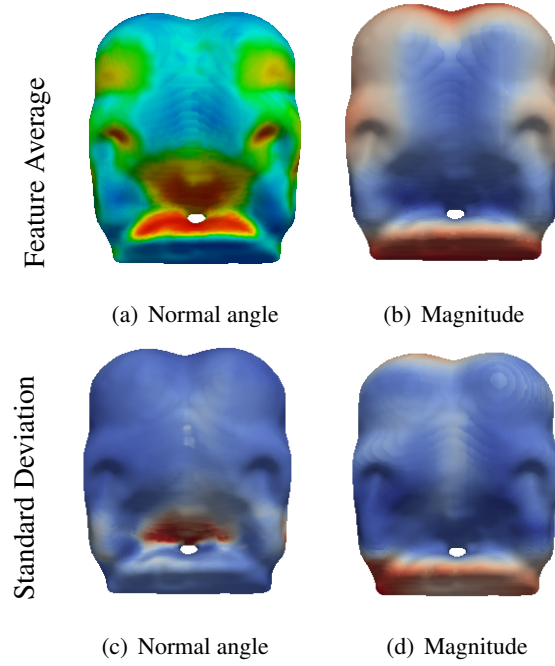
Figure 7.4: Average and standard deviation heat maps of normal angle and magnitude features representing the differences between stage HH 24 and HH 26 embryos.

has one of the largest sample size groups, so the average growth trajectories have more significance than they do for smaller groups. For each query, every average growth trajectory in the database was ranked according to the spatiogram distance to the query, again using equally weighted magnitude and normal angle feature types. The retrievals were then scored using the evaluation score defined in equation 7.1. In this experiment, the relevant retrieval was the correct growth trajectory template.

The average evaluation scores for each developmental stage group are shown in table 7.2, where each entry in the table represents the average score for a growth trajectory template. The overall average evaluation score for the retrievals was 0.040, which is very close to the ideal minimum value of 0 for a perfect retrieval and validates the methods used.

In addition to validation, this type of retrieval provides another useful tool for classification. An unlabeled difference image representing a transition between two unknown developmental stages can be used as a query to find the most similar average growth trajectories. The most similar result will provide a likely candidate for the growth trajectory. The value of the spatiogram distances used

Table 7.2: Average Retrieval Scores for Template Matching for Stage HH 24 Embryos

	HH 19.5	HH 24.5	HH 25	HH 26	Total average
Average score of template retrieval	0	0.035	0.035	0.091	0.040

for the retrieval ranking also provide information about the quality of the match. When the top ranked results have distances which are close to 1, the maximum value of the spatiogram distance metric, there is high confidence that successful matches were found in the database.

7.5 Discussion

The results from the two similarity retrieval experiments motivate the use of the deformable registration features and the MoG spatiogram framework for describing and comparing image differences. Since they were validated on differences due to normal developmental growth, they can now be applied to cases where the differences between images are not well known. One example of this is the differences between images within a developmental group. There can be a large amount of variability within these groups, leading to difficulties in classifying developmental stages. The methods in this chapter can be applied to quantifying the amount and type of variability within age groups and how this relates to the age of the chick embryos within the group. The standard deviation of average growth trajectories is one way this information can be both quantified and visualized. In addition, growth trajectories for a developmental sequence of chick embryos across and within developmental groups can be compared to show progressive changes over time. An example of a compressed developmental sequence is shown in figure 7.5. Progressive changes in the feature distributions can be seen for both the magnitude and normal angle features.
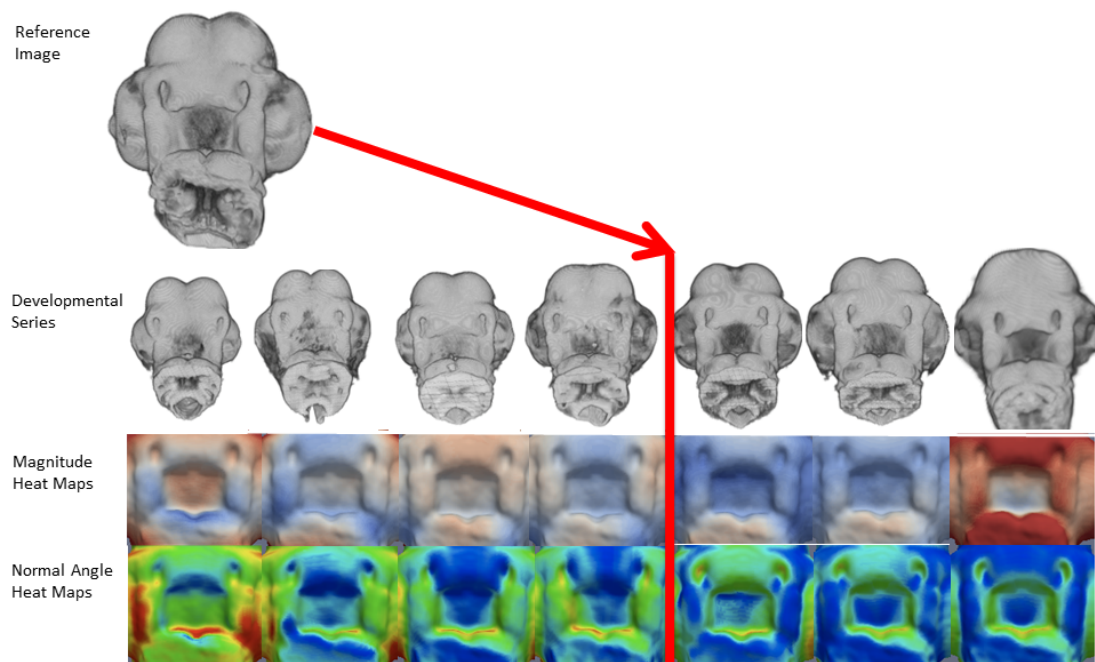
Figure 7.5: Developmental series warped to a template image showing progressive changes in features over time.

Chapter 8

CHICK EMBRYO DEVELOPMENTAL SERIES WITH OROFACIAL CLEFT DEFECT

In the previous chapter, the spatiogram-based framework was demonstrated to be successful in the description of facial morphogenesis in normal chicken embryos. This validation on images with known differences motivates the use of this framework in applications where the differences between images are not well known. This chapter will demonstrate the application of the tools introduced in this work to identify and describe disruptions to normal facial growth caused by genetic or environmental factors.

8.1 Data Set

The data set used for the experiments in this chapter is a set of 20 optical projection tomography (OPT) images of chick embryos from 6 developmental stages that received a transient localized genetic manipulation resulting in a unilateral orofacial cleft. The localized nature of this intervention allowed one side of the embryonic face to develop normally, effectively providing an age matched control for each specimen. An example from each of the 6 developmental stages is shown in figure 8.1. The defect can be seen in the nasal region on the left side of the embryonic face at each stage.

In addition to the complications arising from variability within developmental stage groups discussed in the previous chapter, this data set has further complexity introduced by the clefting defect. First, there is variance in the severity of the clefts. Small differences in the location, delivery, and success of the genetic intervention results in a difference in the expression of the defect. So while the effect of the intervention may be similar within a developmental stage, the severity of that effect can differ. In addition, the asymmetry is known to become more variable as the chick embryos develop. These complications present challenges for analysis, but also help motivate the development of a technique for quantifying and characterizing the variability and differences within the data set. The images from the normal chick embryo data set described in the previous chapter were also included

(a) Stage 1   (b) Stage 2   (c) Stage 3

(d) Stage 4   (e) Stage 5   (f) Stage 5
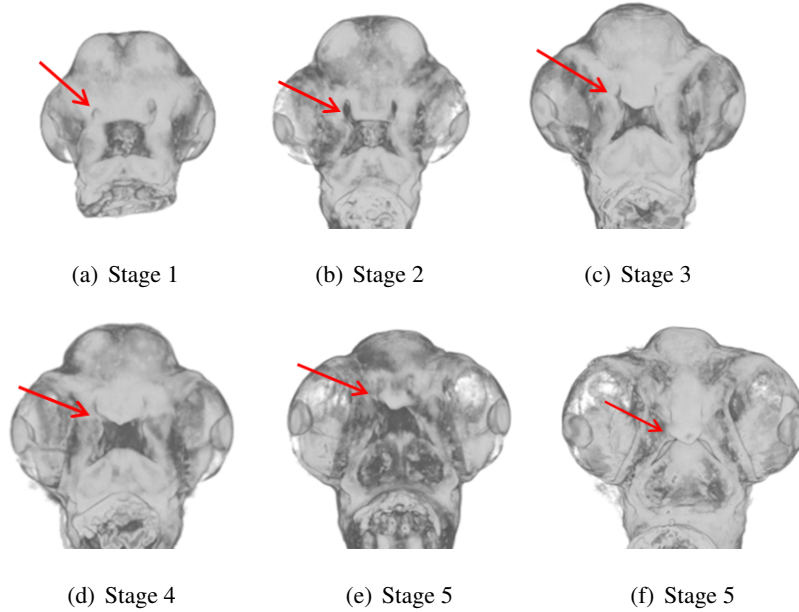
Figure 8.1: Examples of midface development at each developmental (Hamburger and Hamilton [14]; HH) stage in the cleft dataset. The left sides of the images shown are affected by an orofacial cleft and the right side is normal. An arrow points to the affected region in each image.

for experiments requiring controls.

8.2 Symmetry Analysis

Symmetry analysis of the chick embryos is an important way to characterize disruptions in the development of the midface, especially in applications where the shape changes of interest may affect only one side of the face. In this case, comparing the two sides of the face will provide point to point comparisons of normal and abnormal growth while avoiding the challenges related to developmental stage categorization. Two methods of quantifying asymmetry are presented for assessing both bilateral and unilateral facial asymmetry.

8.2.1 Bilateral Asymmetry

The assessment of bilateral asymmetry is important in applications where the shape changes of interest may affect both sides of the face or the impact of the change is not known, so that neither

(a) Original image   (b) Mirrored image   (c) Perfectly symmetric image
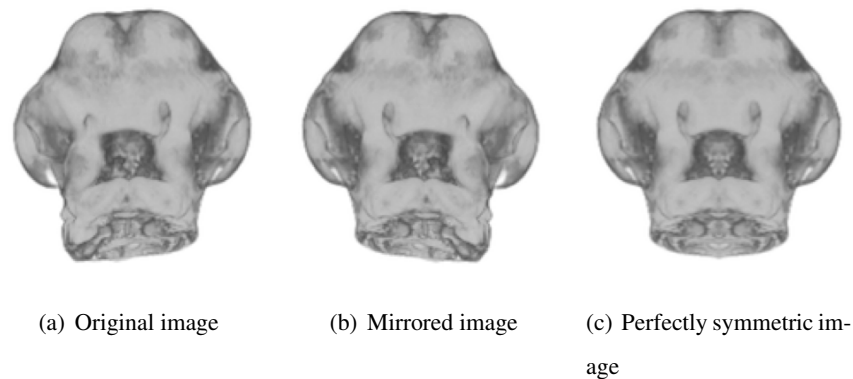
Figure 8.2: Original cleft image and corresponding mirror image and perfectly symmetric image generated by mirroring the right side of the image, that are used to quantify bilateral and unilateral asymmetry.

side of the face can be used as the control. To quantify the bilateral asymmetry of a chick embryo face, both sides are reflected across the midline, determined from the initial alignment landmarks, creating an asymmetric mirrored copy. The mirrored copy is then registered to the midface using the initial registration landmarks. An example of this transformation is shown in figures 8.2(a) and 8.2(b). The bilateral asymmetry is defined by the spatiogram describing the transformation between the original image and its asymmetric, mirrored copy. An example of the visual representation of the bilateral asymmetry feature maps is shown in figure 8.3. The midface region is highlighted in figure 8.3(a) and figures 8.3(b) and 8.3(c) show the corresponding normal angle and magnitude feature maps for the highlighted region.

8.2.2 Unilateral Asymmetry

The assessment of unilateral asymmetry is useful in applications where the shape difference of interest is known to exist only on one side of the specimen, while the other is assumed to be normal. To quantify the difference between the normal and the abnormal side of the face, a perfectly symmetric copy is generated by mirroring the normal side of the face across the midline. An example of this transformation is shown in figures 8.2(a) and 8.2(c). The unilateral asymmetry of the face is then defined by the spatiogram describing the transformation between the original image and its perfectly symmetric copy. Visualizing the features on the affected, left side of the face provides a

(a) Midface region

(b) Normal angle heat map

(c) Magnitude heat map
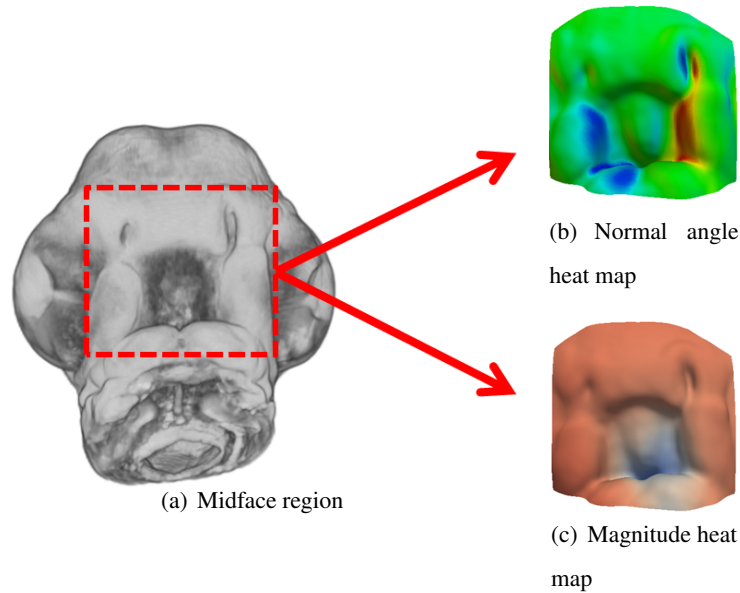
Figure 8.3: Midface of asymmetric image showing features describing difference between the mirrored and original image.

method for viewing the difference from normal development at this stage. An example of this visual representation is shown in figure 8.4. Since unilateral asymmetry exploits the fact that one side of the image is assumed to be normal, this is the preferred measure of asymmetry for the orofacial cleft data set, but both methods of asymmetry will be demonstrated in the experiments in the following sections.

8.3 Classification of Normal and Affected Chick Embryos

The genetic intervention received by the chicks in the orofacial cleft data set is expected to result in an increase in asymmetry, even though the amount of asymmetry is variable and subtle at later developmental stages. Due to this effect, the level of asymmetry can be used to differentiate between normal chick embryos and those that received the genetic intervention. The magnitude of the symmetry transformation is expected to be higher and have more widely distributed values for the specimen in the orofacial cleft data set, as is seen in the magnitude heat maps in figure 8.5. To quantify this, two high-level features from the symmetry magnitude spatiograms were used to rank
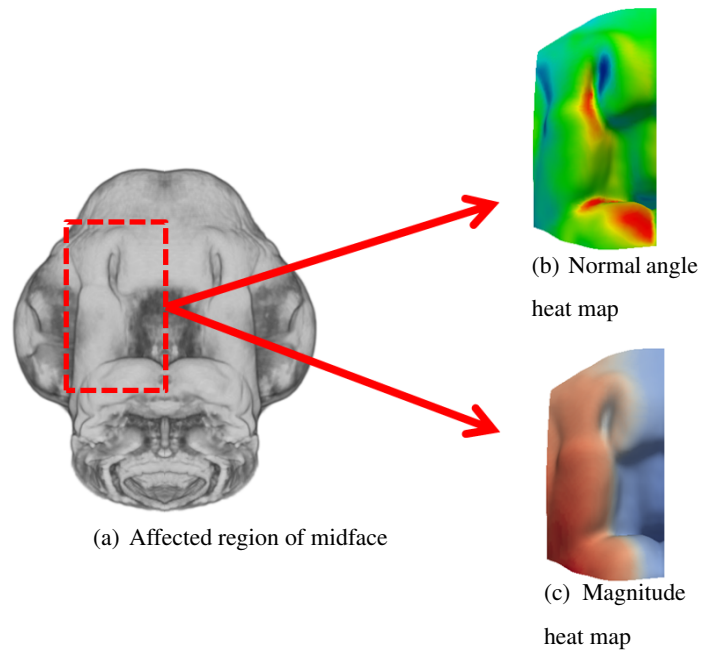
Figure 8.4: Midface of perfectly symmetric image showing features describing difference between the perfectly symmetric and original image for the affected side of the face.

asymmetry:

1. average histogram value,

2. histogram entropy.

These features summarize the magnitude and distribution of the symmetry magnitude transform. Each feature was used to rank the images in the normal and cleft data sets. To rate the performance of the ranking, the average normalized retrieval score used in the previous chapter was used to evaluate the retrieval of cleft images using both features. This score has a value of 0 for a perfect retrieval of the relevant objects in a data set and a value of 1 when all objects in the data set are returned before the relevant objects. The average scores for each feature are shown in table 8.1. The scores are extremely low, demonstrating that even in the case of subtle asymmetry, the cleft images have a higher average magnitude for the symmetry transformation, and the histogram of the magnitude features has a wider distribution of values. These differences allow for successful

(a) Control   (b) Cleft
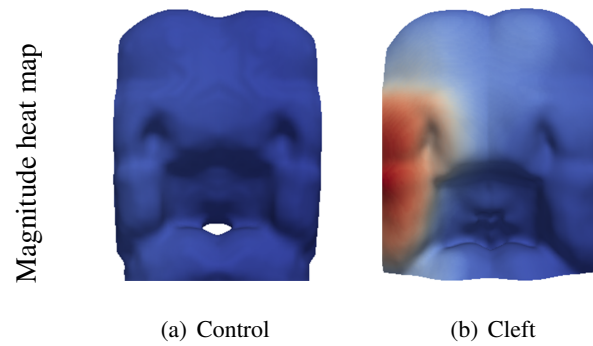
Figure 8.5: Magnitude heat maps showing the difference between the original and perfectly symmetric images for a cleft and control case.

Table 8.1: Classification Performance Using Magnitude Histogram Features

	Average Histogram Value	Histogram Entropy
Retrieval score	0.008	0.017

separation of images from the control and cleft data sets.

8.4 Retrieval of Similar Asymmetry

The features describing the image symmetry can be used to retrieve images from the database of cleft and control images that have similar asymmetry. This is accomplished using a similar method to the similarity retrievals in the previous chapter. One difference image describing the asymmetry of an individual is chosen as a query image. It is expected that a query using a control image would return other control images, and this was shown to be true for unilateral symmetry retrievals, which had a low average score of 0.03. It was also expected that a query using an image from the orofacial cleft data set would return other images at a similar age from this data set. An example of this type of symmetry retrieval using normal angle features is shown in figure 8.6.

For the symmetry queries using an image from the cleft data set, there is one additional factor to

(a) Query   (b) First result   (c) Second result
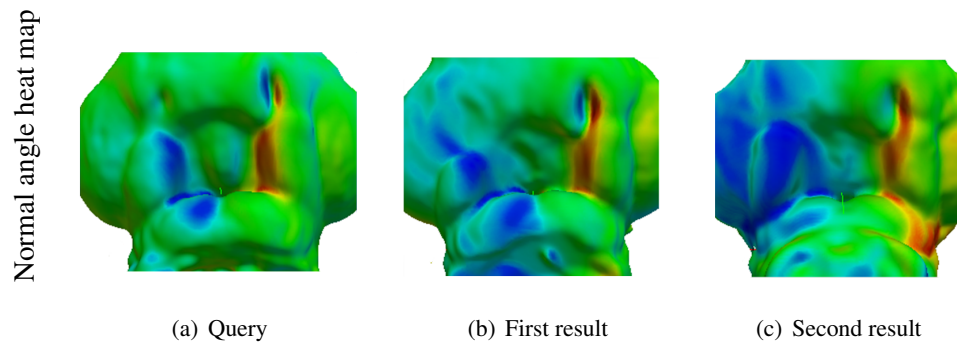
Figure 8.6: Example of symmetry similarity retrieval.

consider concerning similarity of asymmetry. After the genetic intervention, the cleft defect presents as a gap on one side of the face. The shape of this gap changes as the facial prominences grow toward each other. When the chick embryo reaches developmental stage HH 27, fusion of the facial prominences would normally begin to occur. This fusion anchors various parts of the developing facial tissue, enabling subsequent changes in growth and morphology. Small disturbances in the prominences of the midface may delay or completely prevent fusion from occurring on one side. The degree to which the various facial processes are affected can be different from individual to individual. This results in a wider variation in the defects after this point. An example of this is shown for two chick embryos from developmental stage HH 30 in figure 8.7. In figure 8.7(a) the cleft has caused an asymmetric deviation to the left, while in a second individual, in figure 8.7(b) there is an asymmetric deviation to the right.

Due to the variation in presentation of the cleft defects starting at developmental stage HH 27, when fusion would normally occur, it was expected that the symmetry retrieval would produce much better results for the first two stages. This was demonstrated to be true, as shown by the average score for the unilateral symmetry retrieval for each age group in table 8.2. The scores for developmental stages HH 25 and HH 26 are extremely low, indicating successful retrievals and high similarity asymmetry within these groups. The relatively higher scores for subsequent stages indicate less successful retrievals and less similarity of asymmetry within these groups.

While the unilateral symmetry is the preferred method for use with this particular experimental cleft defect, the bilateral symmetry method can also be used for the same retrieval task. The average
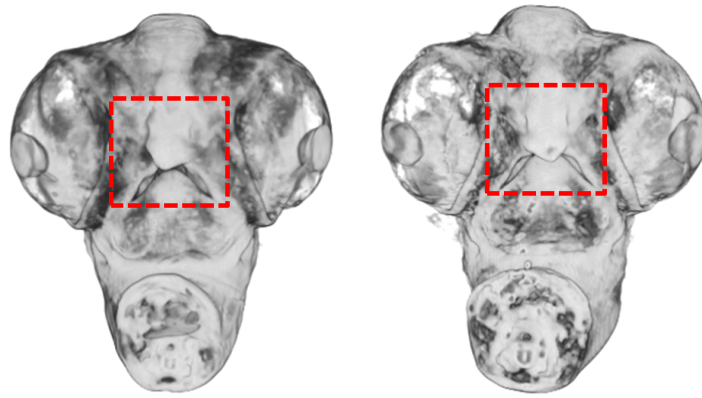
(a) Affected face showing asymmetric deviation to the left

(b) Affected face showing asymmetric deviation to the right

Figure 8.7: Examples of orofacial clefting from developmental stage HH 30 showing variability in expression.

Table 8.2: Retrieval of Similar Unilateral Asymmetry

Developmental Stage	Retrieval Score
HH 25	0.000
HH 26	0.003
HH 27	0.250
HH 28	0.191
HH 29	0.123
HH 30	0.133

Table 8.3: Retrieval of Similar Bilateral Asymmetry

Developmental Stage	Retrieval Score
HH 25	0.050
HH 26	0.022

scores for the retrievals using the bilateral symmetry are very close to those using the unilateral method, and very low scores indicating successful retrievals were achieved for developmental stages HH 25 and HH 26, as shown in table 8.3.

8.5 *Using the Average Neighborhood Difference Feature to Identify Regions of Interest*

The mid-level feature, *average neighborhood difference*, measures feature organization within neighborhoods and can be used to detect regions where the features have high local variance. In this application, regions of high average neighborhood difference represent areas where the symmetry transformation features are changing and can be used to detect areas of significant asymmetry. An example of this is shown in figure 8.8. The normal angle heat map is shown in figure 8.8(a) and the corresponding average neighborhood difference feature map is shown in figure 8.8(b). It can be seen in figure 8.8(b) that the areas of high local variance, shown in red, are the regions most affected by the asymmetry due to the orofacial cleft defect.

Since the orofacial cleft defects are consistent within each of the first two developmental stages, they also show consistency for this mid-level feature. This is shown in figure 8.9, where three average neighborhood difference heat maps are shown from one developmental stage. In each image, it can be seen that the regions most affected by the cleft defect show the highest average neighborhood difference and that these regions are similar across the three images shown.

The regions of high local variation identified by this feature can be used as regions of interest for retrieval. In this experiment, a threshold was set for the average neighborhood difference and spatiograms were generated using only the features in regions with values above this threshold. The threshold was chosen to be 0.8, ten times the mean value of this feature in a normal specimen.

(a) Normal angle heat map

(b) AND of normal angle heat map
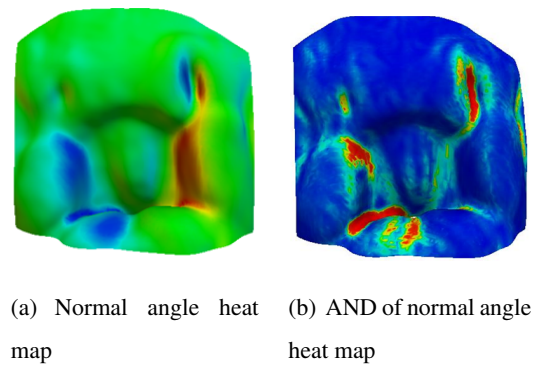
Figure 8.8: Symmetry normal angle heat map and corresponding average neighborhood difference (AND) heat map for a cleft case.

AND heat map

(a) Query

(b) First result

(c) Second result
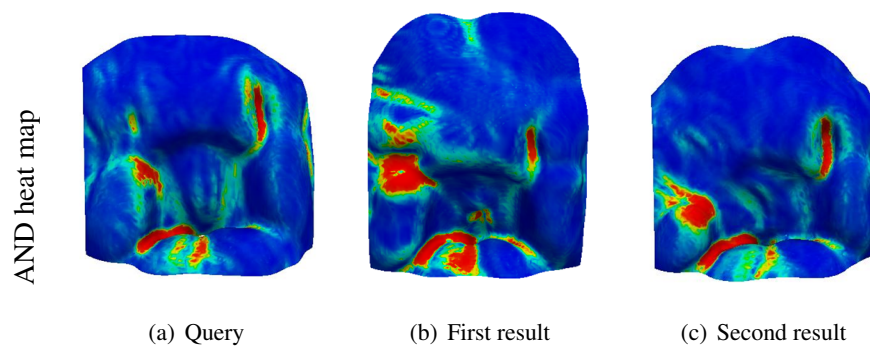
Figure 8.9: Example of average neighborhood difference (AND) of normal angle retrieval for a cleft case.

Table 8.4: Retrieval of Similar Unilateral Asymmetry Using Features from Entropy ROIs

Developmental Stage	Retrieval Score
HH 25	0.000
HH 26	0.017

Table 8.5: Retrieval of Similar Bilateral Asymmetry Using Features from Entropy ROIs

Developmental Stage	Retrieval Score
HH 25	0.035
HH 26	0.014

Each of the specimens in the HH 25 and HH 26 developmental stage groups were used as a query image and the queries were scored by the average normalized rank of the images from the same developmental stage group. The average query score for each developmental stage group is shown in table 8.4. These scores are very low and indicate that this is a very effective way to find regions of significant differences. Comparing these results to the results in table 8.2 shows that the retrieval performance is similar to using the features from the entire midface. Therefore, it can be concluded that the information needed to retrieve images with similar asymmetry was captured by the average neighborhood difference ROIs used in this experiment.

This retrieval experiment was repeated using the bilateral symmetry method to demonstrate that it could achieve similar results. The average retrieval scores in table 8.5 show that the performance of the two methods is similar.

8.6   Average Transition Templates

In the previous chapter, average difference images for transitions between developmental stages were demonstrated to be useful for both template matching and visualization of the spatial distribution of difference features. Here, they are applied to the transitions between abnormal growth stages. Since it is known from the previous experiments that the later developmental stages have high variability in the expression of the midfacial defects, the average transition templates are less representative of the individual transformations than they are for the normal templates. However, the standard deviation maps of the features provide useful information regarding the distribution of the differences in the expression of the orofacial cleft defect within developmental stage groups. Figure 8.10(a) shows the standard deviation heat map for all transitions between developmental stages 25 and 26, before fusion occurs. It can be seen that the region affected by the cleft defect has low values of standard deviation, shown in blue. A compressed sequence of the standard deviation heat maps for transitions between subsequent stages is shown in figure 8.10(b). These images show the highest variance in regions impacted by the cleft defect at these stages, shown in red. This provides a method for both quantifying and visualizing the effect of the variation in the cleft defect expression at these stages.

8.7   Discussion

The experiments in this chapter have shown that the framework presented in this work is useful for quantifying and characterizing the shape changes that occur due to genetic or environmental factors where the differences were not previously known. The tools presented in this work were able to use two different measures of asymmetry to separate specimens with subtle midfacial defects from controls, and to retrieve specimens with similar defects from a database. The methods were also used to quantify and visualize the variation in the orofacial cleft defects that occurred in later developmental stages.
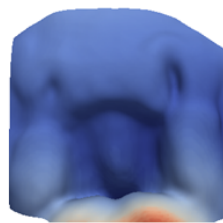
(a) Standard deviation heat map for transition between stages HH 25 and HH 26
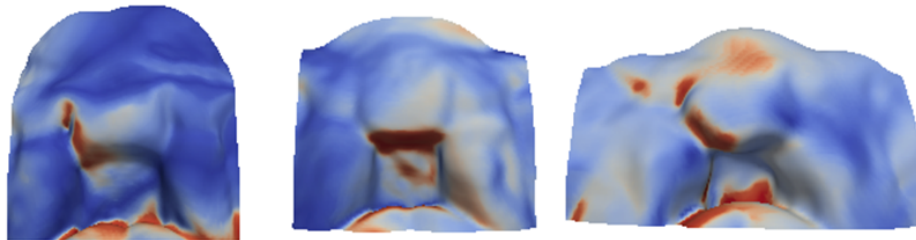
(b) Standard deviation heat maps for later developmental stage transitions

Figure 8.10: Heat maps showing standard deviation of normal angle features for sequential transitions between developmental stages. The growth trajectory between stage HH 25 and HH 26 shows no increase in the cleft region where later stages show high variability.

Chapter 9

CONCLUSION

9.1 Summary

This work develops a landmark-independent deformation-based registration framework that can utilize 3D surface images generated by any multi-dimensional imaging modality. For all datasets, shape differences between developmental stages, treatment groups, or wildtype and mutants can be assessed using regional representations of the dense vector field features. Because of the micron/submicron resolution of the imaging modalities and the voxel sensitivity of the computational algorithm, this method permits quantitative assessment of potentially biologically relevant changes in 3D shape whether natural, caused by mutation, or influenced by non-genetic factors.

The utility and sensitivity of the tools presented in this work were validated on several example data sets. First, the low-level and mid-level deformation field feature extraction methods were applied to synthetic data where the differences between images were due only to known transformations. Regional clustering of the features was able to identify and characterize these known transformations. Next, the deformation vector field features were used to create a tool for assessing the symmetry of bilaterally paired objects. This tool was used to quantify and compare mandibular asymmetry in normal and mutant mice bred to express a midfacial defect similar to craniofacial microsomia, aiding research into the developmental basis of this condition.

A framework based on spatiograms was then introduced to summarize, integrate, and represent the spatial distribution of the features describing image differences used in the previous chapters. This new representation allowed for the definition of a distance measure to quantify the similarity between shape change trajectories. Quantification of the similarity provided a method for similarity retrieval, where results are retrieved from a database based on similarity to the shape change trajectory used as a query. The framework was first validated on normal morphological change over a time series of facial development for chick embryos. The developmental stage labels were used as ground truth to assess the performance of the growth trajectory queries and the retrieval of similar

growth trajectories was shown be highly successful.

After validation on the normal data set, the framework was applied to explore shape differences in a data set containing chick embryos impacted by a local genetic manipulation affecting facial morphogenesis. This demonstrated the utility of these tools for both describing differences from normal growth and quantifying variance within the data set. The features were also shown to be effective at identifying regions of shape change that are scientifically interesting and the performance of the similarity queries was replicated using only features from these regions, validating that they provided an effective summary of the information in the image.

9.1.1 Main Contributions

The main contributions of this work are:

1. a new method for extracting smooth object surface contours from noisy 3D scan data;

2. a new method for extracting regions of important shape change from deformable vector fields;

3. a tool to assess the asymmetry of bilaterally paired structures;

4. the application of this tool to quantify and describe the asymmetry of normal and mutant mice;

5. a new framework for representing morphometric shape changes;

6. a new distance metric for calculating the distance between morphometric shape changes;

7. the application of these methods to assess normal embryonic development;

8. the application of these methods to abnormal embryonic development to characterize the differences from normal development.

9.2 Limitations

This section will discuss the limitations of the methods presented in this paper and how possible sources of error have been quantified and addressed.

9.2.1 Producing closed meshes

This work presents a method for extracting smooth, closed surface contours from problematic and often noisy 3D images. The geodesic contour component of the algorithm balances adherence to the image boundaries and smoothness of the contour to exclude noise (either bright pixels or holes) in the boundary. However, noise in the image boundary is difficult to handle when it is similar in scale to relevant curvatures in the surface of an object, since the level of boundary smoothing necessary to eliminate the noise may also eliminate relevant features. Images with this type of noise can be corrected manually before analysis. If an image contains a large number of these surface discontinuities, manual corrections may be time consuming and not practical.

9.2.2 Landmarks for alignment

The methods in this work present an alternative to using landmark points for shape analysis, but a small number of landmarks are still needed for initial alignment. This alignment is a crucial step in the analysis since it defines the reference frame from which the differences are evaluated. Automatic methods were implemented for initial alignment, but it was found that for some applications minimizing the mean-squared difference between the two images did not produce an alignment that was consistent or biologically meaningful for the regions of interest. While the data sets analyzed do not need to have a large number of reliable landmarks, they are required to have enough for alignment.

The manual identification of initial alignment landmarks is one potential source for error in the analysis. To quantify the amount of this error, an experiment was done in which the landmarks were repeatedly placed on 4 chick embryos over 3 separate days. The average standard deviation for the landmarks was $25\mu m$, which was determined to be reasonable for the data set by the researchers providing the images.

9.2.3 Deformable registration runtime

The runtime of the deformable registration algorithm will greatly depend on the computational power. For a given amount of available computational power, the runtime introduces a constraint on the size of image differences that can be expressed using this method. For high resolution images like those in the chick embryo data set, this runtime can vary between one hour and eight hours,

depending on the scale of the differences between the two images. Image comparisons between objects with very different sizes, such as those where the embryos have a difference in scale of a magnitude of 5 or more, may not be practical to assess due to the long run times that would result. This is not a concern for the applications in this work since there were no size differences that resulted in unreasonable runtimes in the data sets used. If large size differences of this scale were of interest, the resolution of the images could be reduced to mitigate this issue. Alternatively, computational power could be increased.

The parameters of the deformable registration can also affect the runtime of this algorithm. The optimizer's cost function convergence factor determines the minimum reduction in the cost function that is allowed before the algorithm terminates. For the experiments in this work, a typical threshold for moderate accuracy was used to balance need for efficient runtimes with a high level of accuracy. To verify that decreasing this threshold would not effect the results, a subset of the image comparisons were redone with this parameter set to a typical value for high accuracy, resulting in very long runtimes. The queries were then repeated using the features extracted from the high accuracy deformable field. The average retrieval score from the high accuracy features showed a slight increase from 0.030 to 0.035 when compared to the original retrieval score. This validated the use of the threshold for moderate accuracy. If further reductions in runtime were necessary, it is possible that this error threshold could be reduced further without impacting the accuracy of these methods.

9.3 Future Work

The results presented in this paper have opened up several new possible directions for the continuation of this work. The flexibility of the morphometric shape analysis tools makes them ideal for a wide range of applications and imaging modalities, especially for applications where the identification of a large number of consistent landmark points is difficult or impossible. Researchers at SCRI have identified several biological and medical research areas that would benefit from the tool set developed in this work.

(a) Control to control 1   (b) Control to control 2

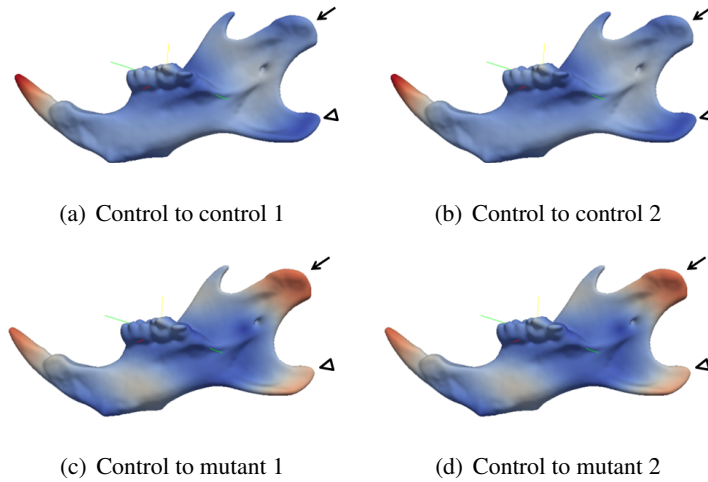(c) Control to mutant 1   (d) Control to mutant 2

Figure 9.1: Magnitude heat maps for the difference between hemi-mandibles from two specimen. Each heat map in the first row shows the difference between two controls and each in the second row shows the difference between a control and a mutant. Differences between the controls and mutants have higher magnitudes, shown in red. The condylar and angular processes are indicated with arrows and arrowheads, respectively.

9.3.1   Comparing Mutant and Normal Mouse Mandibles

In this work, the bilateral symmetry tool was applied to assess asymmetries between paired left and right mandibles from single specimens. In the future, these tools could be used to assess shape differences between mandibles from control and mutant mice, or even digitally dissected hemi-mandibles from clinical CT scans of patients and controls. Figure 9.1 shows an example for this application in mice. Figures 9.1(a) and 9.1(b) each show a magnitude heat map for the difference between two control specimens. Figures 9.1(c) and 9.1(d) each show the magnitude heat map for the difference between a control and a mutant specimen. While all specimens have slightly higher magnitudes in the incisor region due to tooth wear, it can be seen that there is a significantly larger magnitude for difference between the control and mutant specimens in the condylar and anglular processes. This suggests that the magnitude feature and other tools presented here would be able to provide useful information about the differences between the mutant and control groups in this data set.

(a) Control to control 1   (b) Control to control 2   (c) Control to control 3

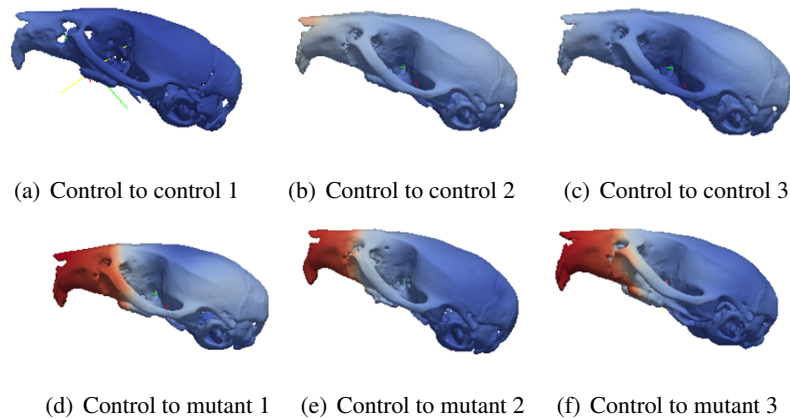(d) Control to mutant 1   (e) Control to mutant 2   (f) Control to mutant 3

Figure 9.2: Magnitude heat maps for the difference between mouse skulls. Each heat map in the first row shows the difference between two controls and each in the second row shows the difference between a mutant and a control. The differences between the controls and mutants have higher magnitude in the snout region, shown in red.

9.3.2 Comparing Mutant and Normal Mouse Skulls

These techniques can also be used to analyze a variety of structures, such as adult mouse skulls. The mouse mandible data set has corresponding skull images for each specimen, and shape differences between the control and mutant skull shapes are expected. An example of the magnitude feature applied to the mouse skull data set is shown in figure 9.2. Figures 9.2(a) through 9.2(c) each show a magnitude feature heat map for a comparison between two control specimens. Figures 9.2(d) through 9.2(e) each show a magnitude feature heat map for a comparisons between a control and a mutant specimen. It can be seen that the differences between the control and mutant specimens again show higher magnitude, this time in the snout region, motivating the use of these tools for further analysis.

9.3.3 Assessing the Effect of Ethanol on Facial Development in a Chick Model

Since the tools in this work were shown to be effective for quantifying and describing both normal chick embryo facial development and facial development impacted by an orofacial cleft defect, they can now be applied to assess a variety of abnormal embryonic growth trajectories. Of particular interest is a data set of control chick embryos and chick embryos exposed to ethanol that are

expected to show subtle and characteristic differences from normal development. Providing quantitative data on these differences will help further the understanding of the effect of maternal alcohol consumption.

9.3.4   *Expanding the Graphical User Interface*

A web-based graphical interface has been developed for the distribution of the tool for assessing symmetry in bilaterally paired structures introduced in chapter 5. In the future, adding the functionality of all the tools described in this work would help to promote their ongoing use and application to new datasets.

BIBLIOGRAPHY

[1] J. Ashburner, C. Hutton, R. Frackowiak, I. Johnsrude, C. Price, and K. Friston. Identifying global anatomical differences: deformation-based morphometry. *Human Brain Mapping*, 6(5-6):348–357, 1998.

[2] Stanley T Birchfield and Sriram Rangarajan. Spatiograms versus histograms for region-based tracking. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 1158–1163. IEEE, 2005.

[3] F.L. Bookstein. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical Image Analysis*, 1(3):225–243, 1997.

[4] R.H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

[5] J. Cao and K.J. Worsley. The detection of local shape changes via the geometry of hotelling's t 2 fields. *Annals of Statistics*, 27(3):925–942, 1999.

[6] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International journal of computer vision*, 22(1):61–79, 1997.

[7] B. Combes and S. Prima. New algorithms to map asymmetries of 3d surfaces. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, pages 17–25, 2008.

[8] Ciarán Ó Conaire, Noel E O'Connor, and Alan F Smeaton. An improved spatiogram similarity measure for robust object localisation. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 1, pages I–1069. IEEE, 2007.

[9] WR Crum, T. Hartkens, and DLG Hill. Non-rigid image registration: theory and practice. *British journal of radiology*, 77(Special Issue 2):S140, 2004.

[10] Tron A Darvann, Nuno V Hermann, Per Larsen, Hildur Ólafsdóttir, Izabella V Hansen, Hanne D Hove, Leif Christensen, Daniel Rueckert, and Sven Kreiborg. Automated quantification and analysis of mandibular asymmetry. In *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, pages 416–419. IEEE, 2010.

[11] A.A. Goldstein. Convex programming in hilbert space. *Bull. Amer. Math. Soc*, 70(5):709–710, 1964.

[12] J.C. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.

[13] B. Hallgrímsson, C.J. Dorval, M.L. Zelditch, and R.Z. German. Craniofacial variability and morphological integration in mice susceptible to cleft lip and palate. *Journal of anatomy*, 205(6):501–517, 2004.

[14] Viktor Hamburger and Howard L Hamilton. A series of normal stages in the development of the chick embryo. *Journal of morphology*, 88(1):49–92, 1951.

[15] L. Ibanez, W. Schroeder, L. Ng, and J. Cates. *The ITK Software Guide*. "Kitware, Inc. ISBN 1-930934-15-7", http://www.itk.org/ItkSoftwareGuide.pdf, second edition, 2005.

[16] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.

[17] E. Kristensen, T.E. Parsons, B. Hallgrímsson, and S.K. Boyd. A novel 3-d image-based morphological method for phenotypic analysis. *Biomedical Engineering, IEEE Transactions on*, 55(12):2826–2831, 2008.

[18] J.L. Lancaster, P.V. Kochunov, P.M. Thompson, A.W. Toga, and P.T. Fox. Asymmetry of the brain surface from deformation field analysis. *Human brain mapping*, 19(2):79–89, 2003.

[19] Damian M Lyons. Sharing landmark information using mixture of gaussian terrain spatiograms. In *Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems*, pages 5603–5608. IEEE Press, 2009.

[20] D. Mattes, D.R. Haynor, H. Vesselle, T.K. Lewellyn, and W. Eubank. Nonrigid multimodality image registration. In *Proceedings of SPIE*, volume 4322, page 1609, 2001.

[21] Joseph G McCarthy, BH Grayson, PJ Coccaro, and D Wood-Smith. Craniofacial microsomia. *Plastic surgery*, 4:3054–3100, 1990.

[22] P.A. Mossey, J. Little, R.G. Munger, M.J. Dixon, and W.C. Shaw. Cleft lip and palate. *The Lancet*, 374(9703):1773–1785, 2009.

[23] Henning Müller, Stéphane Marchand-Maillet, and Thierry Pun. The truth about corel-evaluation in image retrieval. In *Image and Video Retrieval*, pages 38–49. Springer, 2002.

[24] H. Ólafsdóttir, T.A. Darvann, B.K. Ersboll, E. Oubel, NV Hermann, AF Frangi, P. Larsen, CA Perlyn, GM Morriss-Kay, and S. Kreiborg. A craniofacial statistical deformation model of wild-type mice and crouzon mice. In *International Symposium on Medical Imaging*. Citeseer, 2007.

[25] H. Ólafsdóttir, T.A. Darvann, N.V. Hermann, E. Oubel, B.K. Ersboll, A.F. Frangi, P. Larsen, C.A. Perlyn, G.M. Morriss-Kay, and S. Kreiborg. Computational mouse atlases and their application to automatic assessment of craniofacial dysmorphology caused by the crouzon mutation fgfr2c342y. *Journal of Anatomy*, 211(1):37–52, 2007.

[26] H. Ólafsdóttir, M. Hansen, K. Sjostrand, T. Darvann, N. Hermann, E. Oubel, B. Ersboll, R. Larsen, A. Frangi, P. Larsen, et al. Sparse statistical deformation model for the analysis of craniofacial malformations in the crouzon mouse. *Image Analysis*, pages 112–121, 2007.

[27] H. Ólafsdóttir, S. Lanche, T.A. Darvann, N.V. Hermann, R. Larsen, B.K. Ersboll, E. Oubel, A.F. Frangi, P. Larsen, C.A. Perlyn, et al. A point-wise quantification of asymmetry using deformation fields: application to the study of the crouzon mouse model. In *Proceedings of the 10th international conference on Medical image computing and computer-assisted intervention*, pages 452–459. Springer-Verlag, 2007.

[28] N. Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11:285–296, 1975.

[29] T.E. Parsons, E. Kristensen, L. Hornung, V.M. Diewert, S.K. Boyd, R.Z. German, and B. Hallgrímsson. Phenotypic variability and craniofacial dysmorphology: increased shape variance in a mouse model for cleft lip. *Journal of Anatomy*, 212(2):135–143, 2008.

[30] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on pattern analysis and machine intelligence*, 12(7):629–639, 1990.

[31] J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever. Mutual-information-based registration of medical images: a survey. *Medical Imaging, IEEE Transactions on*, 22(8):986–1004, 2003.

[32] PD Polly. Adaptive zones and the pinniped ankle: a 3d quantitative analysis of carnivoran tarsal evolution. *Mammalian Evolutionary Morphology: A Tribute to Frederick S. Szalay. Springer: Dordrecht, The Netherlands*, pages 165–194, 2008.

[33] F.J. Rohlf and D. Slice. Extensions of the procrustes method for the optimal superimposition of landmarks. *Systematic Biology*, 39(1):40, 1990.

[34] C.E. Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[35] M. Sonka and J.M. Fitzpatrick. *Medical image processing and analysis*. SPIE Press, Bellingham, Wash., 2000.

[36] J.P. Thirion and G. Calmon. Deformation analysis to detect and quantify active lesions in three-dimensional medical image sequences. *Medical Imaging, IEEE Transactions on*, 18(5):429–441, 2002.

[37] G. Wollny, M. Tittgemeyer, and F. Kruggel. Segmentation of vector fields by critical point analysis: Application to brain deformation. *Pattern Recognition*, 1:10524, 2002.

[38] N.M. Young, S. Wat, V.M. Diewert, L.W. Browder, and B. Hallgrímsson. Comparative morphometrics of embryonic facial morphogenesis: Implications for cleft-lip etiology. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology*, 290(1):123–139, 2007.

[39] M.L. Zelditch, B.L. Lundrigan, and T. Garland Jr. Developmental regulation of skull morphology i: Ontogenetic dynamics of variance. *Evolution & development*, 6(3):194–206, 2004.

[40] M.L. Zelditch, J. Mezey, H.D. Sheets, B.L. Lundrigan, and T. Garland. Developmental regulation of skull morphology ii: Ontogenetic dynamics of covariance. *Evolution & Development*, 8(1):46–60, 2006.