3D Face Shape Analysis

Student: Adam Raine

Supervisor: Professor A.D. Marshall

Moderator: Dr X. Sun

Introduction

Dysmorphology is the study of abnormalities in human development, particularly when these abnormalities present themselves through congenital malformations of various parts of the structure of the human body. These malformations are often due to rare genetic syndromes, such as Miller-Dieker syndrome, Noonan syndrome, Velocardialfacial syndrome (VCFS), and Williams syndrome (Winter, 1996). As such, many of these dysmorphic syndromes can be diagnosed by looking for these abnormalities, particularly on the face. However, inexperienced geneticists may have trouble making an accurate diagnosis, due to limited experience with certain syndromes, or due to lack of exposure to a sufficient range of ages and ethnicities, in their patients. This problem can be alleviated with sufficient time, experience and training, but ideally we would like that time to be minimal. Thus, finding a way to objectively analyse a face for abnormalities would be useful in assisting diagnosis and training (Hammond *et al,* 2004). One way to do achieve this would be to use visual computing techniques to perform an in-depth analysis based on models from 3D facial scanners. The scanner can create a detailed model of a patient's face, which provides a large amount of information in regards to the shape and size of the face, and various facial features. However, unlike the human brain, computers are not trained to recognise faces. What looks like a nose or a set of lips to us, is just a set of discrete sampled points in 3D space to the computer. Teaching it that a set of points corresponds to a nose or an eye is a challenge in itself. Instead, we could get a computer to analyse our 3D models by processing a large set of data to create a model of an average human face. We would then analyse a particular patient's face and compare it against the average, using machine learning techniques such as Principal Component Analysis (PCA). Hammond *et al,* (2004) managed to distinguish between controls, individuals with Noonan Syndrome, and individuals with VCFS with an accuracy rate upwards of 80%. The ultimate goal of my project is to create a program that can read in a set of landmarked scans of faces, and classify facial features based on what it observes.

Background

The following is a set of briefs on some key terms and concepts that I will be using to build the classifier.

Mesh

At the most basic level, a mesh is simply a collection of points that are connected by a set of lines. The space inside a set of connected points is known as a face. Think of a triangle. A triangle is a set of 3 points joined together by 3 straight lines. Each point, or vertex, is connected to the other vertices by two lines, also known as edges. If we were to take that triangle and attach it to three other triangles, we could form a tetrahedron. A tetrahedron has four vertices, each of which is connected to the other vertices via three edges. The space between three connected points is known as a face, so the tetrahedron has four faces, compared to the triangle's one. This set of vertices can be used to describe a volume, or it could be used to describe a surface. A triangle mesh is simply a set of connected triangles, where each vertex of a triangle is a vertex on the mesh, and each edge is described by the list of faces. For each mesh, we have a 3 x *n* matrix of co-ordinates; each row of 3 describes the horizontal component, followed by the vertical component, followed by the depth component, to form a three-dimensional Cartesian co-ordinate in Euclidean space. The list of faces is a 3 x *m* list of faces, where each row contains 3 numbers in the range of 1 to *n*. Each number describes a position in the list of vertices. A triangle mesh can be used to approximate three-dimensional volume or surface. Even rounded shapes can be described; a sculptor can create curved surfaces by chiselling off smaller and smaller flat sections of rock from a large block, we can create the same effect by simply having a large number of vertices connected by very small, straight edges. Computers can deal with straight lines and points much more easily than complex curves,

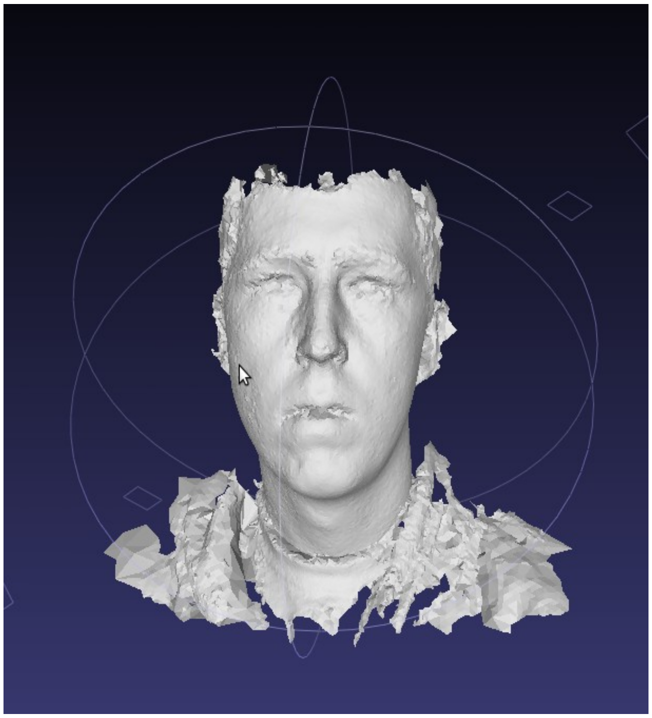

Fig. 1: An image of my own face, taken by a 3D face scanner. Note where the scanner's picked up portions of my clothing; if I were to use this mesh, I would need to cut out data such as that.

so representing a complex 3 dimensional object as a triangle mesh is very efficient and effective.

One complex 3D surface we can represent is the human face. A variety of commercial scanners exist to capture the face and other objects, and such technology is increasingly affordable; Zollhöfer *et al* (2011) describe a process to automatically capture a textured scan of a face using a Kinect camera, a device primarily intended to allow consumers to play video-games using their body.

Registration

Image registration is the process of aligning two images in space so that they share as similar set of co-ordinates as possible. For example, when performing registration on two different face scans would involve lining up the faces so that the eyes, nose and mouth on one mesh are all as close as possible to their positions on the corresponding mesh. The two images should also be as close together as possible. In a sense, the process is like taking two masks of slightly differing shapes, and trying to fit them together as best as possible.

Rigid registration involves transforming a shape or image with six degrees of freedom: translations in the x, y and z axes, and rotations around those axes. The shape itself is not distorted or warped in any way, only moved from one position and orientation to another. If the two images or shapes being registered are different, then the registration will be an approximate fit. In the mask analogy, rigid registration is like taking the masks and aligning them as best as possible, but ensuring that you don't bend or otherwise force either mask into a different shape.

Non-rigid registration is where one shape or image is actively warped to fit another. Rigid registration won't provide a perfect fit if the two objects are different, but non-rigid registration can effectively map one object onto another. This is useful when the structure of the object being moved isn't especially important; for example, non-rigid registration could be used to map the texture of one person's face onto the shape of another.

Registration is important because the face scanner has no context for where a nose or eye is, it simply looks at what is there, and delivers a 3D model of what it sees. People have a variety of different postures and poses, so whereas one person might sit up straight, another might slouch and have their head leaning to one side. Overlaying the surfaces on one another would show two faces in different positions and rotations. To observe the small and subtle differences between faces, it is important that we eliminate large differences and establish a correspondence between the important features on the two surfaces, such as the eyes, nose and mouth. When registered properly, the two faces should overlap.

Landmarks

Landmarks are points on an object that are homologous to all each subject. For example, points on the human face that are common to all humans and found at roughly the same location on each face tested. Points of high curvature, or junctions between object boundaries make good landmarks. Alternatively, easily locatable biological landmarks can be used (Cootes & Taylor, 2004). The landmarks should be easy to identify and place, given that they will generally be placed by experts. However, these points are rare; the inner and outer canthus (the inner and outer corners of each eye) are likely to be present on each subject and easy to spot, but points like the ends of each eyebrow are hard to place exactly, and could have been altered by plucking or other cosmetic enhancements (Hutton *et al*, 2003). There are methods to place landmarks automatically, many of which have been collated by Cootes and Taylor (2004), but the easiest way is to get a trained set of individuals to do the work, especially when working on a model as complex as the human face.

Landmarks are useful because they are good points of reference for registration. If we want to align two faces such that the eyes, nose and mouth are in the same position, it would naturally follow that we use those features as landmarks, register those landmarks, and perform the transformation determined by the registration on the rest of the mesh.

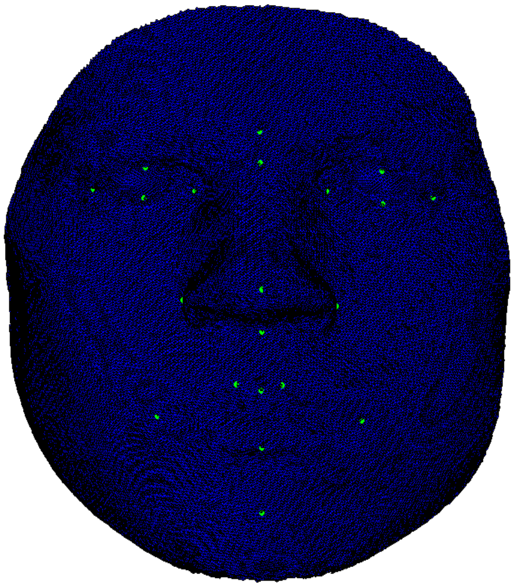

Fig. 2: A mesh in blue with landmarks marked in green

Procrustes analysis

Procrustes analysis is a technique for analysing the distribution of two sets of discrete points, which can form a shape. The idea behind the technique is to minimise the Procrustes distance, which can be defined as $d=\sqrt{(u_1-x_1)^2+(v_1-y_1)^2+(w_1-z_1)^2+\cdots}$, where u,v,w are the horizontal, vertical and depth components of the co-ordinates of one set of points, and x,y,z are the co-ordinates of the other set. The first step in Procrustes analysis is to remove the translational difference between the two sets of points by translating each object's points so that their mean lies at the origin.

Next, the scale difference between the two can be removed by changing the scale of the objects so that their root mean square distance from the points to the origin is 1, where the root mean square distance s is:

$$s=\sqrt{\frac{(x_1-\bar{x})^2+(y_1-\bar{y})^2+(z_1-\bar{z})^2+\cdots}{k}}$$

where k is the number of points in each landmark.

The rotational difference is the next component to eliminate. Fix one of the objects so that it is a reference, and then rotate the other around the origin, so that the sum of squared distances between corresponding points from each object is minimised. This should give an optimum angle for rotation.

Using Procrustes analysis is a good method for rigidly aligning landmarks, which are the discrete set of points described in the process.

We can also use Generalised Procrustes analysis to optimally superimpose more than two sets of landmarks, by aligning them such that the sum of distances from each point to the mean is minimised (Cootes & Taylor, 2004). This technique can be used to a align a large number of sets of landmarks so that they are rigidly aligned as closely as possible, without changing the shape of the object formed by the landmark points. The same transformation can be applied to the associated meshes, and as such, can be aligned as closely as possible without warping the mesh and changing its shape. Once the landmarks are optimally aligned, it is possible to find the mean set of landmarks by taking an average the components of each corresponding point. These mean landmarks are useful, as they are a good baseline to align to, and to perform analysis off. The shape model is essentially a set of constraints on an object off a baseline.

Thin-plate spline warping

Thin plate spline warping is a process of non-rigid transformation of a geometric model. It is useful for non-rigid registration, as it can create a smooth transformation of one surface onto another one, whilst minimising distortion. An analogy would be to take a thin metal plate (the titular thin plate), and have a set of anchor points by which we can push or pull the plate (Bookstein, 1989). If we manipulate the anchor points to move them to a new set of positions, the plate will warp and bend to fit. It will maintain the least bending energy possible, *i.e.* it will bend out of position as little as is possible to fit the new position. If this analogy is applied to warping facial meshes, the mesh is a surface, not unlike a moulded thin plate. The anchor points are the landmarks on that mesh, and we move those anchor points to another set of landmarks. The surface is warped as if it was a thin plate, which means that the surface is distorted, but only to fit the constraints. As the thin-plate spline warp minimises bending energy, distortion in the mesh is also minimised (Hutton *et al*, 2003).

This kind of transformation will be helpful in the registration phase to bring the meshes into close alignment after a rigid transformation determined by Procrustes analysis. Since a rigid transformation can only align two meshes to a certain point, non-rigid transformations are necessary to bring the two as close as possible.

PCA

Principal Component Analysis, or PCA, is a method of reducing the dimensionality of a set of data without losing the variance in the information that is being studied. To do this, the variables in the data are transformed to a new set of variables, the principal components (Jolliffe, 2002). These new variables are uncorrelated: a particular principal component will have some degree of correlation with the original variables, but is entirely independent from all the other principal components. As a result, it is easier to observe the parameters that define how a set of data varies; in this case, how facial structure changes within the shape model. PCA doesn't change the data at all, it simply arranges it in a fashion that makes it easier to see which factors affect the data.

The process to perform PCA is the following, as described by Cootes & Taylor (2004):

1) Calculate mean of data

$$\bar{x}=\frac{1}{s}\sum_{i=1}^{s}x_i$$

2) Calculate covariance of the data

$$S=\frac{1}{s-1}\sum_{i=1}^{s}(x_i-\bar{x})(x_i-\bar{x})^T$$

3) Calculate eigenvectors and eigenvalues λ_i of S (sorted so that $\lambda_i \geq \lambda_{i+1}$

The eigenvectors are the principal components. Processing data along these axes will allow for a much simpler analysis of any data, compared to analysing it against a set of interconnected variables.

Statistical shape model

A statistical shape model is a model of the shape of an object, formed from many examples of that kind of object. The shape of an object is defined by a set of n points which are located at a consistent location from one shape to the next. For example, on a face, one point may be the very tip of the nose, or the corners of the eyes. These points are typically in 2 or 3 dimensions, if we wish to describe a shape in the real world. A shape model is formed by taking a set of a particular kind of object, faces for this project, and locating a set of points on each individual object, such as the locations of biological features. The model describes the manner in which these points can be distributed and still form a valid example of that object. For example, all faces usually have 2 eyes that are roughly level with each other, with a nose that is roughly perpendicular to the axis that the eyes form, and a mouth underneath that which is roughly parallel to the eyes. If we see an example which does not fit these constraints, then the object is either not a face, or not aligned correctly.

To form a shape model, we first align all the points using Generalised Procrustes analysis. This aligns the points to the mean set of points. After that, the set of residuals taken from the distance between each point and its corresponding mean point can be used in Principal Component Analysis to form the model. Since the human face has a high degree of complexity, and not many points that can reliably be used as landmarks for the Procrustes analysis, I will be taking an extra non-rigid alignment step to closely align each mesh with an average, and instead be taking residuals from the difference between points on the mesh, rather than the landmarks. This should result in a much more accurate analysis.

Classification

Classification is a kind of machine learning, and is the process of categorising sets of data. Typically, we first have a training set of data that has been pre-classified. A model of the data is created, and then any new sets of data should be classified based on that model. For example, I might show a computer a number of images of fruit. A successful classifier would be able to tell me what type of fruit it was looking at. It should also be able to reject the object as not being a fruit at all, by putting it into a null class.

For my project, I will be trying to classify different areas of the mouth and lips as described by Wilson *et al* (2012). The data I have has already been classified, so once the shape model has been formed, different variations of the shape need to be assigned to certain categories of lip features.

The approach I will be taking with regards to the classification is Clustering. The basic approach behind clustering is simply delineating sets of points based on a certain criteria, such as distance from the centroid of a cluster, or the density of a cluster. When the PCA has been performed, I will have a number of principal components. Since the data pertains to the whole face, and I'm only really interested in the shape of the mouth, many of the principal components will be of no use to me. I'll need to isolate the components that affect the shape of the mouth first. Then, I need to take the data, pick a particular lip feature, and sort the data by feature type. Each type needs to have its own cluster; the best clustering method will likely need to be found through experimentation. I make a model based on a portion of the available data, and then test the classifier's accuracy based on data I haven't used, since I can compare what the classifier finds against the expert human classification. When observing new items of data to be classified, the mesh should be compared to the average mesh in the same way that the training set was, which should allow for a comparison to the rest of the shape model.

Data

The data I will be using is from the Avon Longitudinal Study of Parents and Children (ALSPAC), also known as Children of the 90s. It is a long term health research project involving the study of over 14000 children as they grow up. I have a selection of untextured, anonymous facial scans from this group, and each scan has been manually landmarked and classified by Caryl Wilson of the School of Dentistry. The landmark points and their location on the human face are as follows:

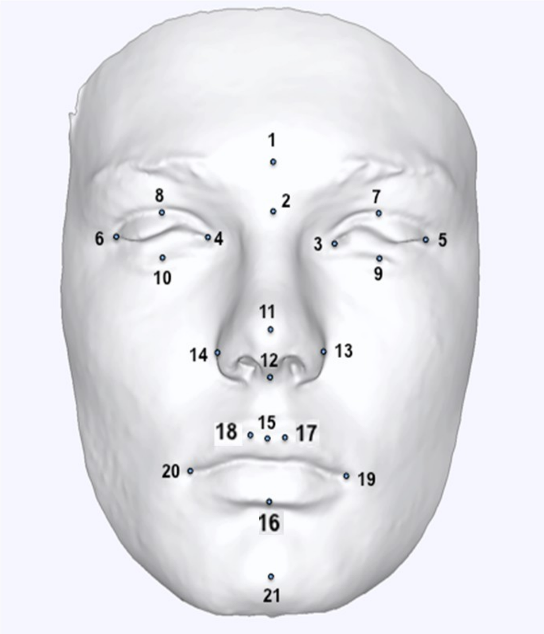

Fig. 3: An illustration of where each landmark lies on a face

Glabella (1) – Most prominent midline point between eyebrows
Nasion (2) – Deepest point of nasal bridge
Endocanthion L/R (3/4) – Inner commissure of the left and right eye fissure
Exocanthion L/R (5/6) - Outer commissure of the left and right eye fissure
Palpebrale superius L/R (7/8) – Superior mid-portion of the free margin of upper left and right eyelids
Palpebrale inferius L/R (9/10) – Inferior mid-portion of the free margin of lower left and right eyelids
Pronasale (11) – Most protruded point of the apex nasi
Subnasale (12) – Mid-point of angle at columella base
Alare L/R (13/14) – Most lateral point on left and right alar contour
Labiale superius (15) – Mid-point of the upper vermillion line
Labiale inferius (16) – Mid-point of the lower vermillion line
Crista philtri L/R (17/18) – Point on left and right elevated margins of the philtrum just above VL
Cheilion L/R (19/20) – point located at left and right labial commissure
Pogonion (21) – Most anterior mid-point of the chin

The classifications of the data are detailed by Wilson *et al* (2012), and include different properties of the philtrum, cupid's bow, upper and lower vermillions, sub-lip and nasolabial angle. Each mesh has a number assigned for each of the different mouth features, which denotes which category that feature falls under.

Approach

The approach I will be following to build the classifier is essentially the one presented by Hutton *et al* (2003), given that it was used to analyse facial morphology by Hammond *et al* with success (2004).

The first step in the process is registration. Each mesh in the data set has a set of landmarks, but these need to be aligned to an average set of landmarks so that there is a common frame of reference. The approach followed here is a Generalised Procrustes Analysis technique broadly similar to the one described by Cootes & Taylor (2004):

1) Translate each set of landmarks so that their centre is at the origin. This puts each set of landmarks is in a roughly similar frame of reference.
2) Choose any set of landmarks as the initial estimate of the mean landmarks. Every other set will align to this set first.
3) Align each set of landmarks to the estimated mean using Procrustes analysis. This will minimise the sum of the distances from each landmark point to its corresponding point on the mean estimate.
4) Once each set of landmarks has been aligned, take a new estimate of the mean.
5) If the mean estimate has not changed significantly (*i.e*. below a certain threshold), then we have convergence, and the we take the mean set of landmarks to be the estimate. Otherwise, go back to step 3.

Now that we have a set of mean landmarks, the rigid registration step is complete.

The next step is to warp each mesh onto the mean landmarks using the thin-plate spline technique. This will closely align each mesh with the mean landmarks, so that we can later reverse the transformation after a correspondence with a base mesh has been established. Once this process has been completed for every mesh, we pick a base mesh, ideally one that has no holes or other glitches where the scanner hasn't quite managed to pick up the face accurately, or where it has picked up extraneous data such as clothing. This will be the mesh that we match every other mesh to.

The correspondence process is fairly simple. Each mesh should be closely aligned with the base mesh, as they have all been warped to fit the mean landmarks. First, we ignore any vertices that are more than a certain distance (the exact value will need to be refined) away from the surface of the base mesh. This is to eliminate any data that is too far from the ideal base mesh, which could be hair or clothing. Next, for each point on the warped mesh, the closest point on the base mesh is found. This establishes a correspondence between the points on the base mesh and the warped mesh.

Since we now have a correspondence, the old landmarks are no longer necessary, and the vertices essentially become the new landmarks. The connectivity of the base mesh is transferred to the warped mesh, and the inverse of the original thin-plate spline warp is performed to move vertices from the warped mesh back to their original positions. The correspondence means that it is much easier to see how every part of the face changes, not just the landmarks. However, it also vastly increases the number of variables we have. Since the vertices can now be treated like landmarks due to the correspondence, the Procrustes algorithm can be applied again to align to align all the surfaces to find a mean mesh, in a similar manner to how the average landmarks were calculated earlier. The average position of each point is calculated, and a new mesh is formed using those average points. Observing how the mesh varies between subjects will be how the statistical shape model is formed, and identifying commonalities between certain subjects will be key to classifying new faces.

Next, Principal component analysis is applied to the data to form the shape model. This will interpret the data so that a number of uncorrelated variables can be used to observe how faces vary within the model. The number of variables should be much less than the number of variables we currently have; each vertex on the mesh is a variable. PCA is applied to the set of residuals of co-ordinate differences between the points on the average face and the points on the processed faces.

Now that the shape model has been obtained, the final step is to use machine learning techniques to teach the program to classify certain facial features. Since the original data from the dental school has already been classified by hand, the problem will be matching up the model with the features that have been classified; the shape of the lips in this case. Since we are only concerned about that area of the face, the biggest task will be isolating changes in lip size and shape from the changes in the rest of the face, that we are not as interested in. Once the factors have been isolated, it should be possible to build a classifier so that, when given a landmarked mesh, the program can compare it the shape model that has been built, and classify the features it finds according to what has previously been observed, ideally with a high degree of accuracy.

Conclusions

Thus far, I have successfully completed Procrustes analysis on sets of test landmarks in order to test the basic alignment process. I have taken this further, and warped their corresponding meshes using thin-plate spline warping so that they align closely. I have also managed to form a dense correspondence between two sets of vertices on the mesh by finding the closest vertex on a base mesh for each vertex on another. As such, the basic framework for forming a shape model is in place; these techniques need to be performed on a large scale form the shape model, as I have hundreds of meshes that can be aligned. The Procrustes analysis I've already done needs to be applied to Generalised Procrustes analysis by finding a mean set of landmarks. This will be one of my first tasks, as it is fairly simple to do and creates a good foundation for the rest of shape model. Then, what I've already done in regards to closely aligning the meshes to another mesh needs to be achieved on a large scale, using as much of the data as I can. The process of warping a mesh, forming a correspondence, and then warping it back is computationally expensive for just 2 meshes; the process takes several seconds at the very least. Thus, performing this on all the data I have is likely to be very time consuming, and should be completed as soon as possible so that the project is not bottlenecked by waiting for that to be finished. After that is done, the PCA process will be the next step, as this gives us the shape model. This work should be completed within the first few weeks of the next semester, to allow for as much time as possible to work on the classifier. The hard part will be analysing the results that the PCA gives me so that correspondences between certain features in the data and certain features around the lips can be identified. All the data I have has been classified already, so I will need to isolate certain lip features and identify which principal components correspond to that feature. Since the classifications I have relate only to the lips, there will likely be a lot of junk data relating to other facial features; I will need to isolate the principal components that affect the mouth, and study those in particular. Using machine learning techniques such as clustering will enable the classifier to delineate between different classes of lip features, and classify a new face based on where it falls in the model. A variety of clustering techniques should be tested and tweaked. The classifier won't be 100% accurate, but the accuracy should be maximised by trying to find alternative methods. By the time of the report, I should be able to deliver a working program that can read in a mesh and associated landmarks, and then classify a face according to what it observes, for as many feature types as possible.

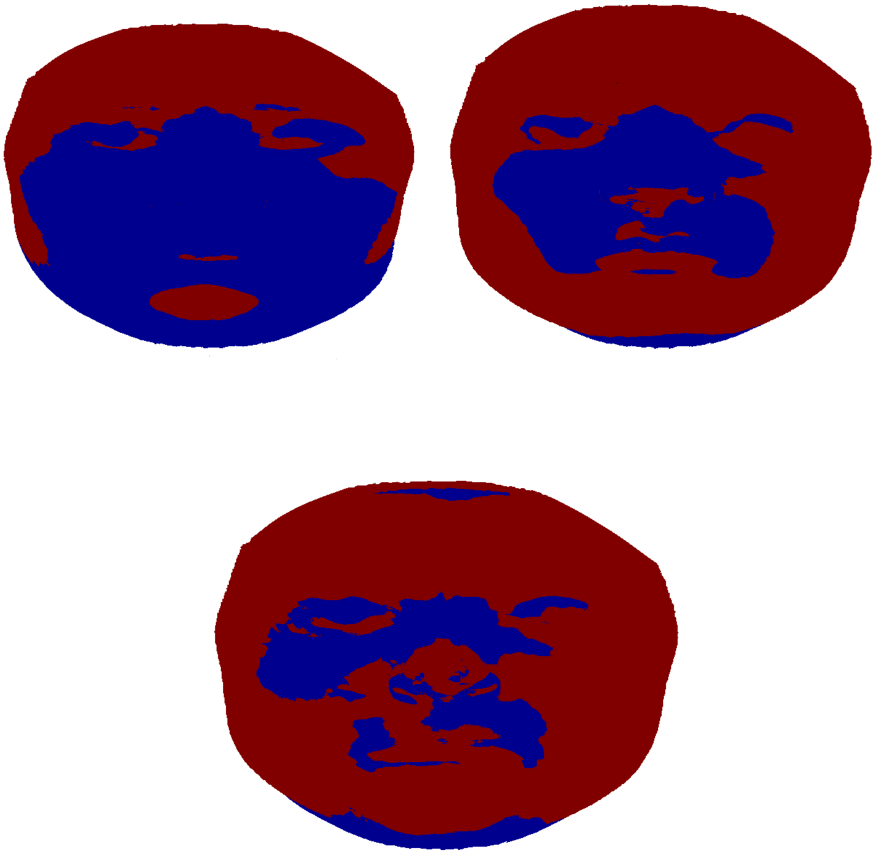

Fig 4: Alignment of 2 meshes. The top left image shows the two meshes overlaid without any alignment. The top right image shows the two meshes aligned with Procrustes analysis, and the bottom image shows the the meshes aligned with thin-plate spline warping.

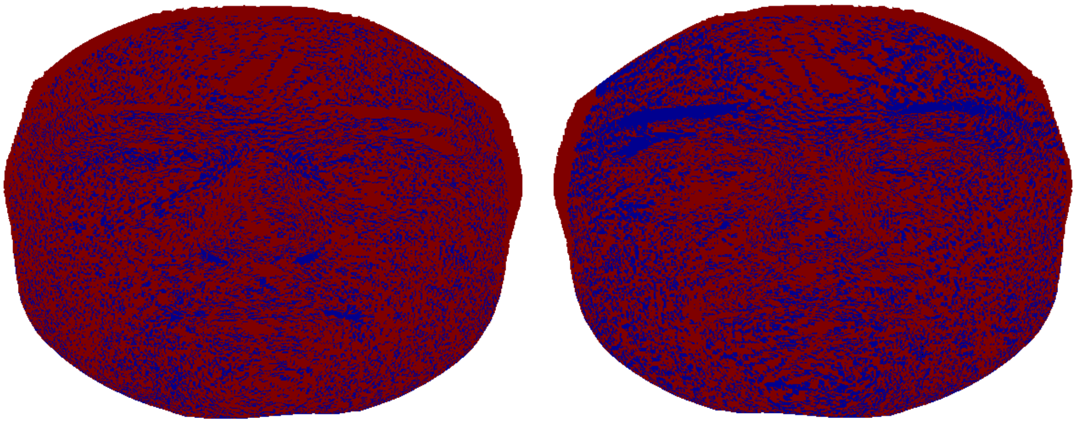

Fig. 5: Alignment of two meshes after the nearest vertices on each mesh have been snapped together. The left shows a view from the front of the meshes, the right shows a view from behind,

References

Bookstein, FL. 1989. Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 11 (6), 567-585.

Bookstein, FL. 1997. Shape and information in medical images: A decade of the morphometric synthesis. Computer Vision and Image Understanding. 66, 97–118.

Cootes, TF & Taylor, CJ. Statistical models of appearance for computer vision. Technical report, Dept of Imaging Science and Biomedical Engineering, University of Manchester, March 2004.

Hammond P, Hutton TJ, Allanson JE, Campbell LE, Hennekam RC, Holden S, Patton MA, Shaw A, Temple IK, Trotter M, Murphy KC & Winter RM. 2004. 3D Analysis of Facial Morphology. *American Journal of Medical Genetics*. 126 (A), 339-348.

Hutton, TJ, Buxton, BF, Hammond, P, Potts, HWW. 2003. Estimating Average Growth Trajectories in Shape-Space Using Kernel Smoothing. *IEEE Transactions on Medical Imaging*. 22 (6), 747-753.

Jolliffe, IT, 2002. *Principal Component Analysis*. 2nd ed. New York: Springer.

Salvi, J, Matabosch, C, Fofi, D, Forest, J. (2007). A review of recent range image registration methods with accuracy evaluation. *Image and Vision Computing*. 25, 578-596.

Wilson C, Playle R, Toma A, Zhurov A, Ness A, Richmond S. 2012. The prevalence of lip vermilion morphological traits in a 15-year-old population. American Journal of Medical Genetics Part A.

Winter RM. 1996. What's in a face?. *Nature Genetics*. 12, 124-129.

Zollhöfer, M., Martinek, M., Greiner, G., Stamminger, M. and Süßmuth, J. (2011), Automatic reconstruction of personalized avatars from 3D face scans. *Computer Animation and Virtual Worlds*, 22, 195–202