Chapter 1

Mathematical Surface Matching of Maps of the Human Torso

Rita Aggarwala[1], Ibrahim Agyemang[2], Erik Andries[3], Melanie Beck[4], Elaine Beltaos[2], Alicia Canton[5], Olivier Dubois[4], Charles Fortin[4], Ana Granados[5], Maria Landry[6], Edmond Lou[7], Matthias Mueck[8], Jeff Orchard[9], Marc Paulhus[10], Adam Webber[11],

Report prepared by Jeff Orchard[12] and Adam Webber[13]

1.1 Introduction

Scoliosis is a fairly common condition that usually appears during adolescence. It involves mild to severe deformations of the spine. Doctors and patients are interested in measuring the progression of the disease. Internal changes in spine angles and twists can be seen from X-rays, but surface (appearance) changes are often of primary concern to adolescent patients. Measuring the amount of surface change can be a difficult problem. Natural growth, changing body shape, and progression of the disease all factor into forming the shape of the patient's torso. Separating these effects, and extracting the change due to scoliosis alone, is challenging. In an effort to quantify the changes due to scoliosis, Capital Health acquires three-dimensional (3D) surface scans of a subject's back every 3 to 6 months using a laser scanner [1]. The problem addressed

[1]University of Calgary
[2]University of Alberta
[3]University of New Mexico
[4]McGill University
[5]University of Washington
[6]University of British Columbia
[7]Capital Health
[8]University of Toronto
[9]Simon Fraser University
[10]Pacific Institute for the Mathematical Sciences
[11]University of Victoria
[12]jjo@sfu.ca
[13]awebber@math.uvic.ca

in this paper is how best to use this data to obtain useful measures for both the doctor and patient.

There are two separate approaches to extract useful information from the data. The first involves making direct measurements on each individual scan and comparing these measurements over time. The second approach is to compare successive images of the back, noting the differences between them. The first approach is simpler, and some measurements of this type have already been implemented by Capital Health. However, these measures are extracted from a 2-dimensional (2D) representation of the data. It is hoped that measures that make use of the 3D data may be more informative. The second approach involves aligning two successive 3D back surfaces to each other. Capital Health currently implements this alignment operation manually [1], and thus measurements taken from these alignments are not completely reproducible due to human error.

Two main issues were taken into account in this project: the reproducibility of the results, and stringent time constraints. In particular, user influence should be removed from each step of the data processing, and results should be obtained within three minutes of acquiring the scan.

The report begins with a description of the data collection, followed by a description of the data processing required to align two back surfaces. A section is devoted to calculating the cosmetic score, a measure of deformity of the back. The paper concludes with a few suggestions for improvements on data collection and use.

1.2   Experimental design and data collection

Patients with scoliosis are examined every 3 to 6 months. Laser surface scanning is a part of this examination, and the entire procedure takes only a few minutes.

Before the scan takes place, several steps are followed. Patients are dressed in a gown and their shoes are removed. The patient is asked to stand normally within a standard frame apparatus, their arms extended to 90° holding onto two poles. The frame has four adjustable guides: one on each hip, and one aligned with the front of each shoulder. The function of the guides is simply to assist the patient in holding the same position during the scan. Although the guides are not used as a measuring device, the nurse records the guide settings.

The laser scanner is approximately 2 metres away from the patient. Once the patient is positioned in the frame, the nurse adjusts the field of view of the camera by selecting one of three zoom settings. The normal setting is 0.72, with 0.6 and 0.8 being used occasionally. To minimize patient motion due to breathing, the patient is asked to take a deep breath and hold it during the scan. The scan takes 0.6 seconds, so patient movement within a single scan is minimal.

For some scans, the nurse places 5 fiducial markers on the patient's back before the patient is placed in the frame: one on the middle of the back of the neck, one on each scapula (at the inferior tip of the shoulder blade), and one on each lower back dimple (the slight indentations on either side of the lumbar spine). Since the markers are placed on easily-identifiable physiological landmarks, we expect their placement to be consistent.

Each scan results in the collection of two different sets of data. The first is a TIFF image,

a digitized photograph of the patient's back. This image contains colour information as well as position information. The second set of data is a cloud of 3D data points (approximately 200×200) that corresponds directly to the TIFF image. Although the scanner is designed to detect the exposed back and arms of the patient, it also tends to detect several other objects within the scanner's range. Automatically stripping the raw data of these extraneous structures was one of the main challenges faced by our group.

1.3   Cropping

The first processing step on the data is to select the portion of data that is relevant for analysis. Each back image (here "image" refers to a set of 3D coordinates) not only includes points describing the surface of the patient's back, but also some points corresponding to the frame, the patient's head, clothing, etc. These points are to be removed by a *cropping* procedure described below.

Our cropping procedure involves only the first two coordinates of each point. As such, it is a 2D cropping procedure. We will denote the first two coordinates by x and y, respectively. Figure 1.1 illustrates an example of an initial (uncropped) image viewed in the xy-plane.
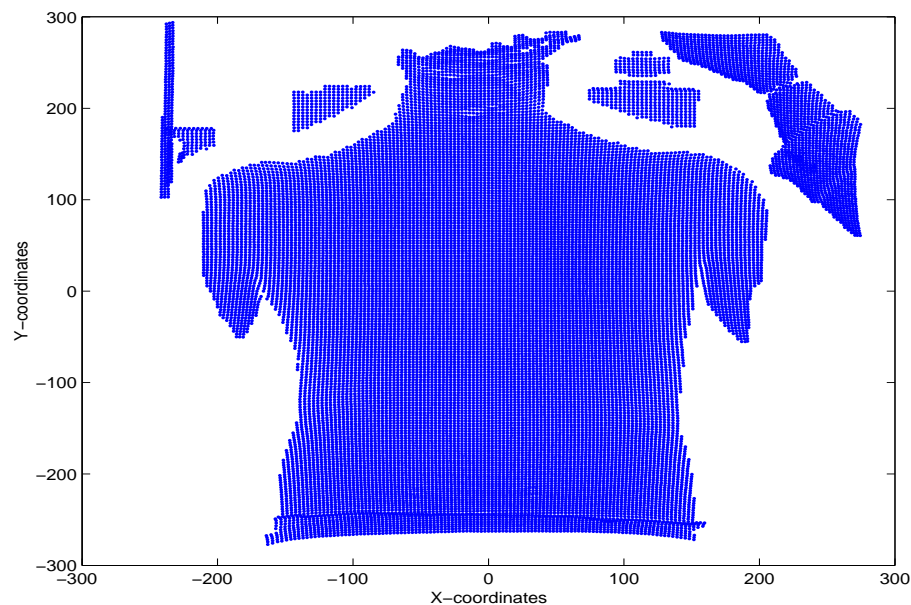

Figure 1.1: Surface data before cropping. This broken 3D surface is projected onto the xy-plane.

The first step of the cropping algorithm is to eliminate the clouds of points which are not connected to the body of the patient. Two points are not connected if there is a large enough space between them where there is no data point. Hence, we need to define what *large enough* means. Let $Y_{\max}$ be the largest Y-coordinate among the points in the data set. Similarly, define $Y_{\min}$, $X_{\max}$ and $X_{\min}$. Now define

$$\epsilon_x := \frac{X_{\max} - X_{\min}}{130} \quad \text{and} \quad \epsilon_y := \frac{Y_{\max} - Y_{\min}}{130} \ .$$

We will say that two sets of points are disconnected from each other if there is an empty rectangle of size $\epsilon_x \times \epsilon_y$ between them.

Although the image may not always be centered, the axis $x = 0$ always runs through the back. We search for a point (x, y) which satisfies

$$0 \leq x \leq \epsilon_x \ , \ Y_{\min} \leq y \leq Y_{\min} + \epsilon_y \ .$$

Once we find one, we look to the right of this rectangle to see if we can find a point (x, y) which satisfies

$$\epsilon_x \leq x \leq 2\epsilon_x \ , \ Y_{\min} \leq y \leq Y_{\min} + \epsilon_y \ .$$

We continue moving to the right by steps of size ϵ_x until we find a rectangle containing no data points. When this occurs, we eliminate all the points which are further to the right of this empty rectangle. The y-value is incremented by ϵ_y, and the procedure is repeated on the next line (i.e. $Y_{\min} + \epsilon_y \leq y \leq Y_{\min} + 2\epsilon_y$). Lines are processed one at a time until $Y_{\max}$ is reached. Then, we repeat all the previous steps, this time moving to the left of $x = 0$, i.e for the points satisfying $x \leq 0$. Figure 1.2 shows the dataset from figure 1.1 after cropping.
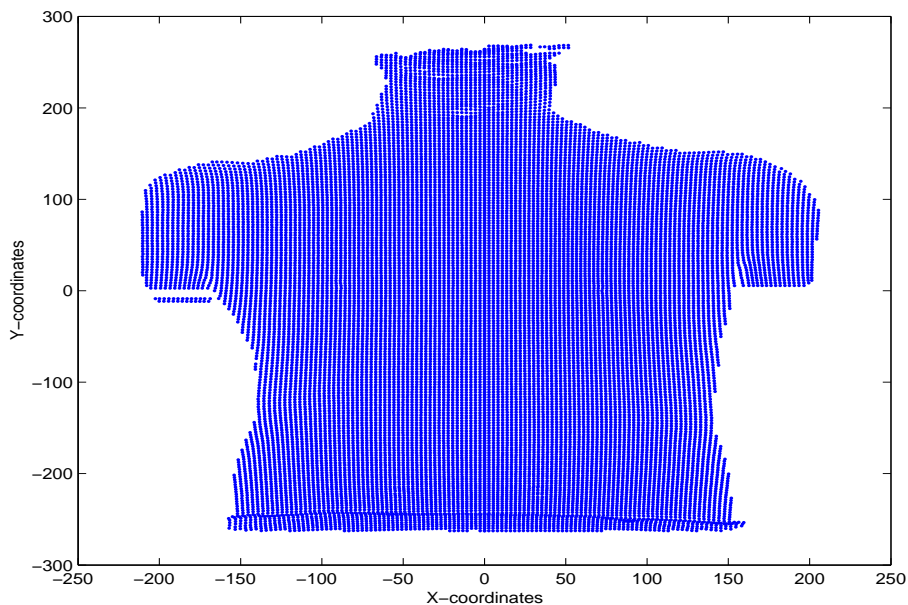

Figure 1.2: Surface data after the first cropping stage. The remaining piece is a single connected component.

At this point, we consider the dataset to be a connected cloud of points. However, it still contains points corresponding to the neck of the patient. To eliminate these points, we simply clear all rows that are shorter than $1/3$ of the width at the shoulders. This finalizes our cropping procedure. Figure 1.3 illustrates the final result for our example.
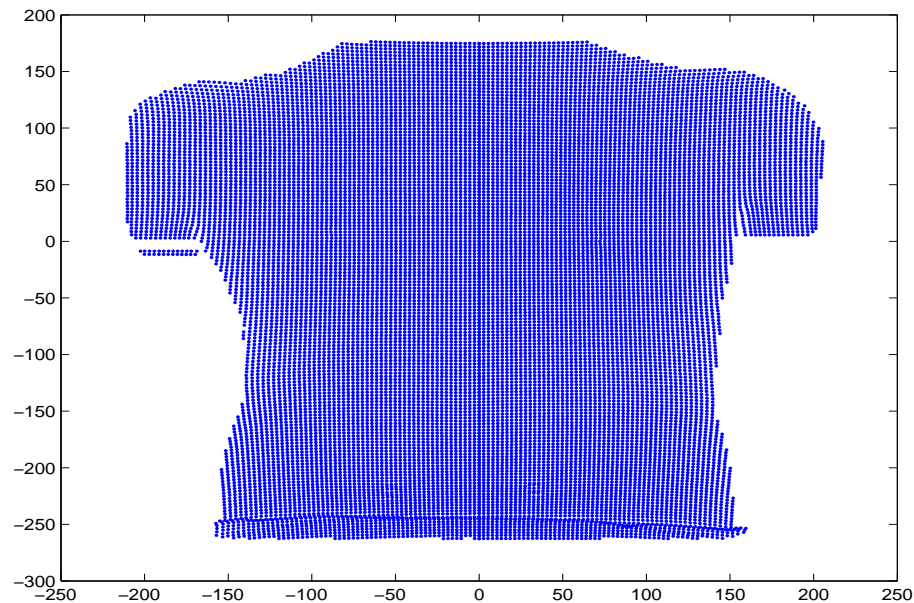
Figure 1.3: Surface data after the final cropping stage. The protruding neck is no longer present.

1.4 Initial Surface Alignment

Recall that our analysis of the progression of scoliosis involves the comparison between two scans acquired at different times, typically 3 to 6 months apart. Once the two images are cropped, we need to align them to allow direct comparison. A good initial guess is essential to get convergence of the alignment fine-tuning phase (see section 1.5). In each alignment phase, we seek an $\mathbb{R}^3 \to \mathbb{R}^3$ coordinate transformation that best aligns one image to the other. To simplify the problem, we consider only a subset of all possible affine maps, along with translations. This simplification is valid because the degrees of freedom for patient motion are limited by the effects of normal growth, the positional frame, and the nature of scoliosis. The set of affine maps that we will consider for the initial alignment are those that are a composition of rotations in the xy-plane (about the z-axis), dilatations for changes in growth and width (scaling along the y-axis and x-axis respectively), and translations in directions x, y and z.

Keeping one of the images fixed, the second is transformed via this affine map to best fit the first image. To establish an initial estimate of the appropriate affine map, we can use the coordinates of corresponding points from each of the two images. That is, each location on the patient's back is sampled in both scans. However, the coordinates of this point may be different in each scan. The affine map is chosen to best match these coordinates.

Since the patient is standing upright and positioned in a fixed frame, it can be assumed that the rotation in the xy-plane is very small, and therefore can be approximated by a y-axis shear in the direction of the x-axis. This simplification enables one to represent the affine map (using

homogeneous coordinates) with the matrix,

$$M = \begin{pmatrix} \nu_1 & \tilde{a} & 0 & t_1 \\ 0 & \nu_2 & 0 & t_2 \\ 0 & 0 & 1 & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where ν_1 and ν_2 are the x and y scaling parameters, t_1, t_2 and t_3 are translation parameters, and $\tilde{a} = a\nu_1$, where a is the shear coefficient.

The system of linear equations given by the matrix M can be rewritten as

$$\begin{pmatrix} \nu_1 & \tilde{a} & 0 \\ 0 & \nu_2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}.$$

Since the unknown parameters are $\nu_1, \nu_2, \tilde{a}, t_1, t_2$ and t_3, we transform this system into

$$\begin{pmatrix} x & 0 & y & 1 & 0 & 0 \\ 0 & y & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \tilde{a} \\ t_1 \\ t_2 \\ t_3 \end{pmatrix} = \begin{pmatrix} x' \\ y' \\ z' - z \end{pmatrix}.$$

We have six unknowns, so we need at least six equations. Since the third row of the matrix does not depend on (x, y, z), it can appear only once (otherwise, a singular system will result). Hence, three reference points are needed to estimate all the motion parameters. If we denote by (x_1, y_1, z_1), (x_2, y_2, z_2) and (x_3, y_3, z_3) the coordinates of these three points in the first scan, and (x'_1, y'_1, z'_1), (x'_2, y'_2, z'_2) and (x'_3, y'_3, z'_3) the coordinates of the three corresponding points in the second scan, we obtain the following system:

$$\begin{pmatrix} x_1 & 0 & y_1 & 1 & 0 & 0 \\ x_2 & 0 & y_2 & 1 & 0 & 0 \\ x_3 & 0 & y_3 & 1 & 0 & 0 \\ 0 & y_1 & 0 & 0 & 1 & 0 \\ 0 & y_2 & 0 & 0 & 1 & 0 \\ 0 & y_3 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \tilde{a} \\ t_1 \\ t_2 \\ t_3 \end{pmatrix} = \begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \\ y'_1 \\ y'_2 \\ y'_3 \\ z'_1 - z_1 \end{pmatrix}.$$

This system is over-determined with seven equations for six unknowns. However, Matlab automatically returns the linear least squares solution. Thus we obtain directly the six parameters that give the best initial fit.

To get a good initial guess, it is very important that the three points be located far apart. One option for locating corresponding points in a pair of scans is to locate the fiducial markers. A Matlab program was developed to automatically detect the position of the markers located on the dimples. The markers used for the scans available to us were (bolt) washers, taped onto the
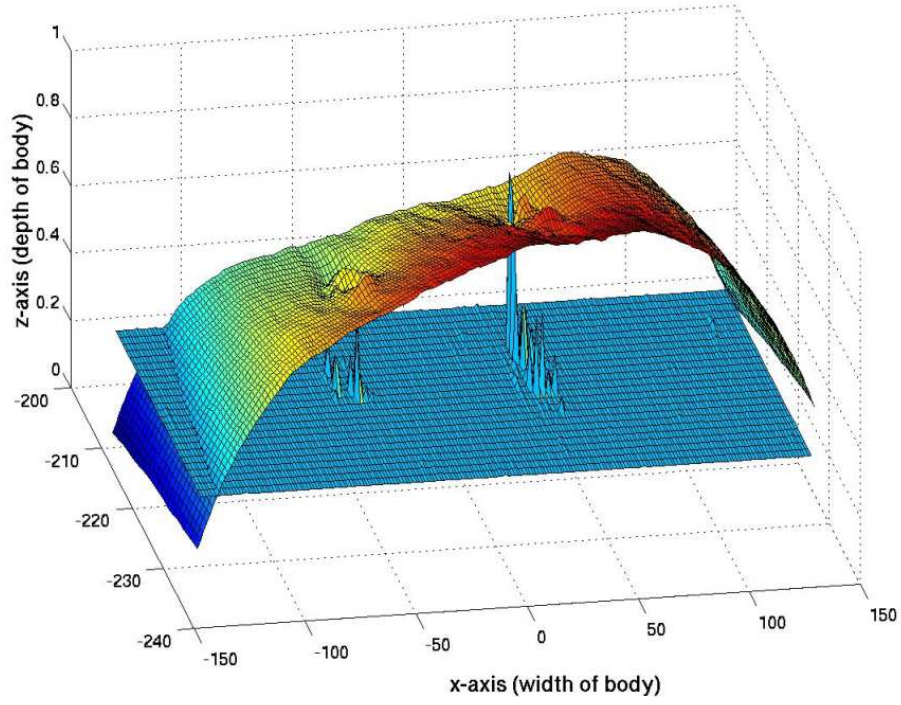
Figure 1.4: Automatic marker detection on the lower back. The spikes in the nearly-flat surface indicate the location of the markers on the back surface (also shown).

patient's skin. These markers are difficult to detect in the 3D surface by visual inspection since they appear as minuscule Heaviside steps in the z direction. However, since they are attached in regions where the torso surface is relatively smooth, they can be detected easily by applying a two-dimensional second-order finite difference stencil,

$$\Delta z(x, y) = z(x - h, y) + z(x + h, y) + z(x, y - h) + z(x, y + h) - 4z(x, y) \ ,$$

to the regions on the torso where the markers are likely to be found. The first difference eliminates a linear trend, while the second difference eliminates a quadratic trend in the data. Since the washers manifest themselves in the 3D cloud of data as annular Heaviside functions,

$$f(x, y) = H\left(r_2 - (x - x_0)^2 - (y - y_0)^2\right) - H\left(r_1 - (x - x_0)^2 - (y - y_0)^2\right) \ ,$$

the second-order differencing eliminates the local quadratic curvature of the torso but does not eliminate the washer-induced Heaviside *roughness*. Hence, by plotting $|\Delta z(x, y)|$ across the relevant regions, the presence of the markers is indicated by large spikes in curvature amplitude. Figure 1.4 shows a portion of the lower back surface, along with $|\Delta z(x, y)|$.

Although the above method for finding corresponding points appears to be effective, points were selected manually for this preliminary study. Two points were selected in the outer part of the shoulders, and one point was selected in the middle of the lower back (between the two dimples).

To illustrate the initial fitting, figure 1.5 shows the initial alignment of two surfaces for the same patient, acquired a few months apart.
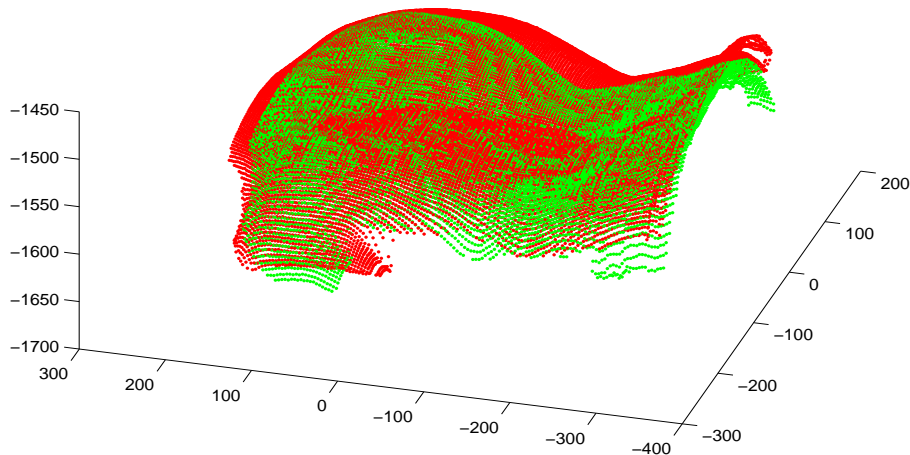
Figure 1.5: Initial alignment of two scans acquired a few months apart. The red surface was aligned to the green surface using the 3-point method outlined in section 1.4.

1.5   Final Surface Alignment

Fine-tuning of the alignment between the two back surfaces is done by adjusting the alignment parameters to minimize a given cost function. Such automatic surface matching has been demonstrated on teeth [2]. Some implementation decisions were made based on the feasibility of performing the optimization on the computers available to us during the workshop.

The automatic re-alignment (registration) method consists of 3 main parts: transformation, cost function, and optimization scheme.

1.5.1   Transformation

The human body is capable of moving into many different positions. The progression of scoliosis accounts for some particular kinds of motion. However, the objective of registration is to remove the effects of body position that are not due to scoliosis. The manual re-alignment process was used to correct a certain subset of typical mis-alignments [1]. In particular, manual alignment allowed the operator to scale, rotate and translate the surfaces. Once the scale factors, rotations and translations are chosen, applying the transformation to a cloud of 3D data points is trivial. However, comparing two clouds of data points that are not sampled at the same locations is quite difficult. Initially, we compared the surfaces by resampling one surface onto the same grid as the other using Matlab's `griddata` function. The `griddata` function performs linear interpolation using a Delaunay triangulation. This method proved to be too slow, since the automatic optimization method required many such comparisons.

Instead of treating the mesh as a general 3D cloud of data points, we converted it to a gray-level image. Such an image uses gray-level to indicate the height of the surface over

the corresponding (x, y) point. Since the surface has a bounded xy domain, a value of zero is assumed anywhere that the surface was not defined. One of the main differences between representing the data as an image instead of a point cloud is that the point cloud data need not be a function, since it can have two points with the same (x, y) coordinates, but different z values. An image can only represent a function. This does not appear to be a major restriction since the original data is range data (which has to be a function), and the alignment rotations tend to be rather small. The advantage to using the image representation is that there are very efficient resampling techniques for regularly gridded image data.

Similar to the subset of affine maps used in section 1.4, the set of transformations used in the final alignment algorithm has 6 parameters (although they are not the same 6 parameters as in section 1.4): translation along the x- and y-axes, scaling along the x- and y-axes, rotation about the z-axis, and rotation about the x-axis. The rotation about the z-axis, the translations and the stretches are all simple 2D image manipulations. However, the rotation about the x-axis is not. Since we expect the x-axis rotation to be quite small, we can approximate it with a shear in the z-direction along the y-axis. This operation is equivalent to adding a y-gradient to the image. All of these operations are essentially 2D, and can be done quickly using built-in Matlab routines.

In this 2D image-based transformation framework, we still use `griddata` to obtain the initial image representation from the cloud of points. All subsequent manipulations are done on the 2D image. One issue, however, is that `griddata` assumes the xy-domain of the surface of interest is the convex hull of all the points. This is a natural assumption to make in general, but does not apply in this context. Figure 1.6 shows the image representation of a torso surface. Notice that the region inside the line joining the waist and the outer edge of the arms is not treated as background, even though it is clearly not part of the back (see figure 1.3 for comparison). This is a property of the interpolation method used in `griddata`. Further pre-processing to insert background points in that region may solve the problem. However, for this study, the full convex hull was used.

1.5.2   Cost Function

A measure of goodness-of-fit must be established in order to perform automatic alignment. For this study, we chose to use the sum of squared residuals between the image representations of the two back surfaces. There are many other cost functions that can be utilized, but none others were implemented for this report.

It should be noted that the sum of squares cost function is sensitive to outliers. The original range data was a cloud of points with z-values between 1400 and 1700. Using these values directly in the gray-level images caused the background to have a very different value from the foreground. As a result, non-overlapping regions between the convex hulls of the two images resulted in extreme outliers, causing very large errors. Thus, the fit was dominated by the amount of overlap. Since the background value of zero was chosen arbitrarily, choosing a different background value could reduce this effect. Equivalently, translating both surfaces so that their minimum z-value is mapped to zero vastly reduces the drastic impact of the distant background. This practice was adopted for our study.
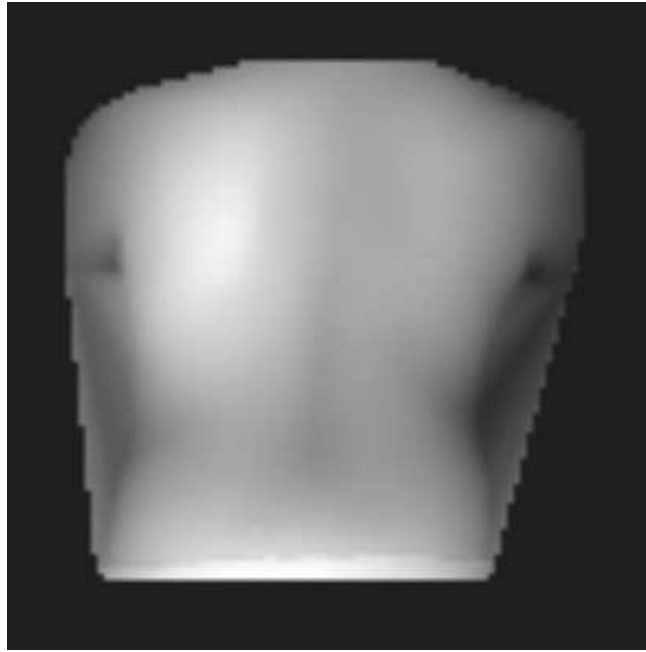
Figure 1.6: Image representation of back surface data. The image has been slightly enhanced to show the nonzero regions between the arms and the waist.

1.5.3   Optimization

The procedure for finding the optimal set of 6 motion parameters was off-loaded to Matlab's optimization toolbox. In a first attempt, we fed the cost function to the unconstrained minimization procedure, `fminsearch`. However, because of the number of parameters, the optimization was very slow and did not always converge to the expected minimum.

To restrict the search, it is easy to find bounds for the motion parameters, and so we used the routine `fmincon` instead. Since we have already performed an initial alignment, the optimal value for each parameter in the final adjustment should be very small. For example, we required that the translations be less than half the range of the data in each direction. The rotations should be limited by 45 degrees. As for the shear corresponding to the forward bending of the patient, we have derived bounds by assuming that the depth of each point should not change by more than the z-range of the data. These bounds are quite loose and could be made tighter.

Figure 1.7 shows the difference between the two back images before and after final alignment. The degree of misalignment is noticeably reduced after the final alignment phase.

1.6   Cosmetic Score

Although the severity of cosmetic deformation may not always be highly correlated with the severity of spinal deformation, quantization of the cosmetic severity is sought by patients making adjustments to living in their society. Investigation into the relationship between the cosmetic deformation and the spinal deformation is not a focus of the present study, though it may prove
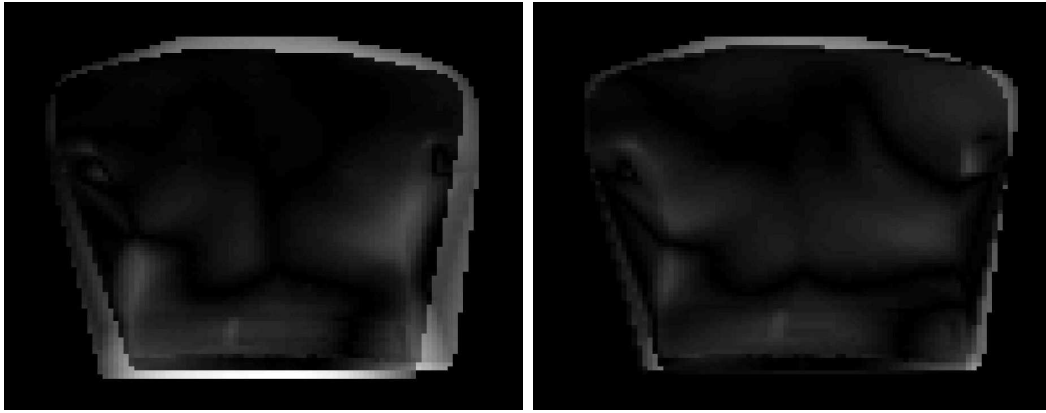
Figure 1.7: Difference images before (left) and after (right) final alignment. The lighter regions show areas of misalignment.

to be useful in future work.

A single, dimensionless number known as the *2-dimensional cosmetic score* is currently calculated as a weighted average of:

- the (scaled) angle between a line joining the two scapula bones (where markers are placed) and a line parallel to the ground

- the (scaled) difference between the angles that the right and left shoulder silhouettes make with a line parallel to the ground

- the waist asymmetry index calculated as a function of several different distances measured from markers to a vertical center line (also determined from markers).

A high cosmetic score is associated with more severe cosmetic deformation.

The dimensionless property of the cosmetic score is appealing, as it is thought to be invariant to an individual's growth over time and positioning in the frame apparatus, and thus the surface alignment issues discussed in previous sections are not of concern here. The simple score is based on the measurement of seven locators on the back, and can be automated easily using, for example, reflective markers that can be detected once cropping has been completed.

The current 2D cosmetic score does not take into account twisting or related depth information available through the 3D laser scan. Generalizing the 2D cosmetic score to three dimensions, while preserving the invariance property of the score, is simple. Possible measurements that may be included in the calculation of a 3D cosmetic score are:

- the twist angle between the line connecting the scapula markers and the line connecting shoulders

- the twist angle between the line connecting the scapula markers and the line connecting dimples in the lower back

- the twist angle between the line connecting shoulders and the line connecting dimples in the lower back.

Since marker position can be obtained in three dimensional Euclidean space, these angles can be measured easily and incorporated into a generalized 3D cosmetic score. Furthermore, automated marker localization will make the calculation of this score fast and reproducible. Additional measures of importance may be considered with clinical input from medical experts familiar with the disease.

Perhaps a more interesting and revealing exploration is a comparison between the current 2D score and a generalized 3D score. A "perfect" 2D score can obviously be associated with an imperfect 3D score. Also of interest is be the predictive ability of the final cosmetic score. Determining the formula for the score requires not only determining which measurements are most important in describing the severity of cosmetic deformation, but also verification by an experiment that can incorporate errors associated with operator variation, repeated measurements over time, and so on. Finally, an investigation comparing the cosmetic scores for scoliosis patients with normal (scoliosis-free) and post-operative individuals will be informative and important in determining the utility of the ideas currently in use as well as those being presented here.

1.7 Conclusions

We have shown that, in principle, the data processing required to crop and align the surfaces can be done automatically, and within the 3-minute time window, as required. However, there are simple procedural modifications that can make this processing more robust and practical.

The scanner zoom settings are not standardized between visits. In the future, the scanner settings should be recorded and then reused in successive visits, minimizing the need for the alignment algorithm to accommodate for large scale changes.

The markers should be designed so that they are more easily detected in the surface data. Markers with a deeper cross-section, or with reflective properties could be used to create a very distinct and noticeable bump on the 3D surface. Once these markers can be found automatically, efficiently, and robustly, they should be used to automate the initial alignment procedure.

In addition to what was demonstrated in this study, further study using more degrees of freedom in the motion model may yield important results. A particular subset of motion parameters (eg. a twisting along the spine, or y-axis) could be used to model the deformation caused by scoliosis, thereby enabling the alignment algorithm to quantify the progression of scoliosis with motion parameters. It should be noted, however, that the particular types of motion used in this study could be implemented on the 2D image representation of the back data. A deformation such as twisting could be more difficult to implement on the 2D images.

Data from various physiological features and the location of markers might be useful for calculating a 3D cosmetic score that represents the severity of observable deformity. This 3D score will likely contain more specific information than the currently used 2D score. A study should be conducted to evaluate various forms of the 3D score and determine what particular mixture of measurements best encapsulates the deformity, and is least susceptible to measurement/operator error.

Bibliography

[1] D. C. Berg, D. L. Hill, J. V. Raso, E. Lou, T. Church, M. J. Moreau, J. K. Mahood. Using three-dimensional difference maps to assess changes in scoliotic deformities. *Medical & Biological Engineering & Computing*, Vol. 40, February 2002, pp. 1-6.

[2] H. L. Mitchell, R. G. Chadwick. Mathematical shape matching as a tool in tooth wear assessment - development and conduct. *Journal of Oral Rehabilitation*, Vol. 25, 1998, pp. 921-928.