3D Reconstruction of the Human Jaw from A Sequence of Images

Computer Vision and Image Processing Laboratory
University of Louisville, Department of Electrical Engineering
Louisville, KY 40292
E-mail:mohamed@cairo.spd.louisville.edu, Phone:(502)-852-6130, Fax:(502)852-1580

Abstract—A novel approach is proposed to obtain a record of the patient's occlusion using computer vision. Data acquisition is obtained using intra-oral video cameras. The technique utilizes shape from shading to extract 3D information from 2D views of the jaw, and a novel technique for 3D data registration using genetic algorithms. The resulting 3D model can be used for diagnosis, treatment planning, and implant purposes. The overall purpose of this research is to develop a model-based vision system for orthodontics to replace traditional approaches. This system will be flexible, accurate, and will reduce the cost of orthodontic treatments.

Keywords—Image Sequence Analysis, Shape Representation, Registration.

I. INTRODUCTION

ORTHODONTIC treatment involves the application of force systems to teeth over time to correct malocclusions. In order to evaluate tooth movement progress, the orthodontist monitors this movement by means of visual inspection, intra-oral measurements, fabrication of plastic models, photographs, and radiographs. This procedure is both time consuming and costly. Also repeated acquisition of radiographs may result in undesired side effects. Obtaining a plaster model of the jaws is a complex operation for the orthodontist, and is very unpleasant experience for the patient and may not provide all the details of the jaw.

Recently, efforts have been devoted to computerize diagnosis in orthodontics. Bernard et al. [2] developed an expert system for orthodontic measurements, however the cephalometric measurements that are fed to the expert system are still acquired manually from the analysis of X-rays and plaster models. In [1], a computer-vision technique was developed for the acquisition and processing of 3D profiles of dental imprints which still requires a plaster model as an input to their system.

The main contribution of this paper is the development of a complete 3D model of the human jaw using combination of shape from shading and a novel registration approach using genetic algorithms.

This paper is organized as follows: Section II presents an overview of our approach to obtain jaw models. Section III describes the application of Shape From Shading to extract 3D information from each jaw image. Section IV is devoted to the registration problem. In this section, a new registration technique using genetic algorithms is presented. Conclusions and future directions are presented in section V.

II. SYSTEM OVERVIEW

As shown in Figure 1, our approach to reconstruct the human jaw consists of five stages. The first stage is data acquisition. Using a very small CCD camera placed inside the oral cavity, a set of overlapping images \{I_j \mid j = 1,2,..,N\} of different parts of the jaw are acquired such that \( \bigcup_{j=1}^{N} I_j \) covers the whole jaw. The setup for acquiring such images is illustrated in Figure 2.

These images are then preprocessed in the second stage to reduce noise and to sharpen edges. This step consists of lowpass filtering and histogram modification. A set of \( N \) 3D models, \{S_j \mid j = 1,2,..,N\}, are then computed using shape from shading. Each one of these models represents a 3D surface description of a part of the jaw. This step is described in section III.

This work was supported in part by grants from the Whitaker Foundation and the NSF (ESC-9505674).
Once these \( N \) models are obtained, a registration technique is used to align these different views together so one model of the whole jaw can be calculated. In this system, we developed a novel registration technique using genetic algorithms as a search engine. The details of this method is presented in section IV. The final stage is to transform this model into patches of free form surfaces using a triangulation technique. This step is used to convert the surface representation of the jaw to STL format which is a standard input for rapid prototyping machines. This can enable us to obtain a rapid prototype model for better visualization.

III. THREE-DIMENSIONAL SHAPE FROM SHADING

The process of recovering surface orientation from gray level shading, Shape From Shading (SFS), was first defined by Horn [3] as follows: Given an intensity image of a continuous surface with constant, known reflectance and illumination, recover the shape of that surface. As shown in Figure 3, suppose that a point is at a distance \( Z \) from the image plane and suppose that orthographic projection is used to map the point onto the image plane. In camera coordinates, the point is at position \((x, y, Z)\). Consider a point nearby in the image plane at position \((x + \delta x, y + \delta y)\). The depth of the point will be \( Z + \delta Z \). Let \( Z = Z(x, y) \). Taking Taylor series expansion of \( Z \) about the point \((x, y, Z)\), the elemental change in the depth \( Z \) can be expressed as follows:

\[
\delta Z \approx \frac{\partial Z}{\partial x} \delta x + \frac{\partial Z}{\partial y} \delta y
\]

Equation (1) relates the change in depth \( \delta Z \) of the point to the change in image plane coordinates \( \delta x \) and \( \delta y \). The size of the partial derivatives of \( Z \) with respect to \( x \) and \( y \) are related to the amount of tilt in the tangent plane at the scene surface corresponding to point \((x, y, Z)\). These partial derivatives are called surface gradients. The gradient vector \((p, q)\) is given by:

\[
p = \frac{\partial Z}{\partial x} \quad \text{and} \quad q = \frac{\partial Z}{\partial y}
\]

The surface normal to a surface patch is related to the gradient by:

\[
n = (p, q, 1)
\]

which simply states that the amount of displacement in \( x \) and \( y \) corresponding to a unit change in depth \( Z \) is \( p \) and \( q \), respectively. The unit surface normal is obtained by dividing the surface normal by its length:

\[
N = \hat{n} = \frac{n}{|n|} = \frac{(p, q, 1)}{\sqrt{1 + p^2 + q^2}}
\]

By assuming that surface patches are homogeneous and uniformly lit by distant light sources, the brightness \( E(x, y) \) seen at the image plane often depends only on the orientation of the surface. This dependence of brightness on surface orientation can be represented as a function \( R(\cdot) \) defined on the Gaussian sphere. Thus, we can formulate the Shape from Shading problem as finding a solution to the brightness equation:

\[
E(x, y) = R(p, q, L)
\]

where \( R(p, q, L) \) is the surface reflectance map and \( L \) is the illuminant direction.

Pentland [9] introduced a method of estimating illuminant direction, \( L \), from the distribution of image derivatives as a function of image direction. The method works by assuming a statistically uniform distribution of the surface orientations, then performing a maximum-likelihood analysis to estimate the cosine variation in image gradient magnitude induced by the directionality of the illuminant.

A. SFS Using Pentland’s Linear Approximation

Two general classes of algorithms have been developed: local algorithms, which attempt to estimate shape from local variations in image intensity and global algorithms which can be further divided into global minimization approaches and global propagation approaches. The global propagation approach attempts to propagate information across a shaded surface starting from points with known surface orientation (singular points), while in global minimization a solution is obtained by minimizing an energy function \[14\]. Among the local algorithms, we present in this section the one developed by Pentland [13] because of its speed and accuracy.

Over a small region we can always approximate the reflectance map by a linear function of the partial derivatives \((p, q)\). This approximation can be determined by taking a first-order Taylor series expansion of \( R(p, q) \) about the central point \((p_0, q_0)\), to obtain

\[
E(x, y) \approx R(p_0, q_0) + (p - p_0) \frac{\partial R(p, q)}{\partial p} \bigg|_{p=p_0,q=q_0} + (q - q_0) \frac{\partial R(p, q)}{\partial q} \bigg|_{p=p_0,q=q_0}
\]

\[
= k_1 + k_2 \frac{p - p_0}{p_0} + k_3 \frac{q - q_0}{q_0}
\]

For Lambertian reflectance function \( k_1 = \cos \tau, k_2 = \cos \tau \sin \sigma, k_3 = \sin \tau \sin \sigma \), where \( \tau \) and \( \sigma \) are the tilt and the slant of the illuminant respectively. This
linear approximation of the reflectance function becomes accurate over a larger area as the illuminant becomes more oblique, and over a smaller area as the illuminant moves closer to the viewer’s direction.

Equation (6) may be transformed into the Fourier domain in order to obtain a convenient and efficient solution. \( k_1 \) is a d.c. term and can be ignored.

\[
\mathcal{F}_E(f, \theta) = k_2 \mathcal{F}_p(f, \theta) + k_3 \mathcal{F}_q(f, \theta) \quad (7)
\]

where \( f \) is the radial frequency and \( \theta \) is the orientation. Since \( p \) and \( q \) are the partial derivatives of the surface height \( Z \), their Fourier transforms are simply

\[
\mathcal{F}_p(f, \theta) = 2\pi \cos \theta \, e^{j\frac{2\pi}{Z}} \mathcal{F}_s(f, \theta) \quad (8)
\]

\[
\mathcal{F}_q(f, \theta) = 2\pi \sin \theta \, e^{j\frac{2\pi}{Z}} \mathcal{F}_s(f, \theta) \quad (9)
\]

Applying the previous equations into equation 6 and substituting for the values of \( k_1 \) and \( k_2 \) we obtain

\[
\mathcal{F}_E(f, \theta) = j2\pi f \sin \sigma (\cos \tau \cos \theta + \sin \tau \sin \theta) \mathcal{F}_s(f, \theta) \quad (10)
\]

where \( j \) is the imaginary axis of the complex domain.

\[
= j2\pi f \sin \sigma \cos(\tau - \theta) \mathcal{F}_s(f, \theta) \quad (11)
\]

let \( \cos(\tau - \theta) = S_d \) then

\[
\mathcal{F}_E(f, \theta) = j2\pi f S_d \sin \sigma \mathcal{F}_s(f, \theta). \quad (12)
\]

\[
\mathcal{F}_s(f, \theta) = \frac{1}{j2\pi f S_d \sin \sigma} \mathcal{F}_E(f, \theta). \quad (13)
\]

This algorithm gives a non-iterative, closed-form solution using the Fourier transform. The problem lies in the linear approximation of the reflectance map when the non-linear terms are large [14] [15].

Figure 4 presents the application of this technique to extract depth information from different 2D jaw images.

Once the 3D information for a series of 2D images \( \{S_j \mid j = 1, 2, ..., N\} \) is obtained, 3D data registration is applied to combine and align these data sets into one model. In the following section, we will present a new technique for data registration using genetic algorithms.

IV. DATA REGISTRATION

Registration is an integral part of computer and robot vision systems. There are three main applications: object recognition, visual navigation, and inspection/validation. The objective of object recognition is to match observed data with a previously stored model representing objects of interest. In visual navigation, the objective is to match data observed in a dynamic scene at different instants in time to be able to recover object motions and to achieve scene interpretation [24]. The objective of inspection/validation is to track data as it changes over time, such as monitoring tumor growth in a patient, in the case of inspection, and to measure on a point by point basis the accuracy of a process, such as comparing the specifications of a manufactured part with the theoretical Computer Aided Design (CAD) model specifications for that part, in the case of validation. The orthodontic application would be another example of using registration for both inspection and validation.

Registration is the process of aligning two data sets for the purposes of calculating a best fit measure between the two sets of data. There are \( 2N + 1 \) degrees of freedom in \( N \)-dimensional space. There are \( N \) rotational degrees of freedom about each axis and \( N \) translational degrees of freedom in the direction of each axis. The other degree of freedom is the result of scaling. Thus it is necessary for the registration process to correct for misalignments in all the degrees of freedom and then calculate an error measure, which is a calculation of the dissimilarity between the experimental and model data sets. The error measure used in this paper is the mean square error. The result of the registration process is a direct transformation from the experimental data set to the model data set and an error measure between the rotated and translated data sets.

There are several approaches to achieve surface registration depending on the application. The methods can be broken down into two main categories: primitive-based and surface-based approaches. The primitive-based approach relies on feature extraction of primitives, such as special points [18], contours [23], and surface patches [22], together with some differential properties invariant to rigid transformation such as Gaussian curvature. In the surface-based approach, the data is considered to be a surface and the goal is to calculate the transformation by minimizing a criterion relating the distance between the two surfaces. Some examples of the surface-based approach include a correlation method [19], and a differential method [20]. One of the main disadvantages of the
above techniques is their computation complexity.

To solve this problem, we present a new primitive-based registration algorithm using genetic algorithms. Our matching criterion function is similar to the one proposed by Zhang [24], except that we use a genetic algorithm as a search engine to speed up computation. The details of this approach is presented in the next subsections.

A. Point-based registration

A parametric 3-D shape $S$, either a curve segment or a surface, is a vector function $x : [a,b] \rightarrow \mathbb{R}^3$ for curves where $a$ and $b$ are scalars and $x : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ for surfaces. Both curve and surface data sets are usually in the form of, or can be easily converted to, a set of 3-D points, thus both data sets can be treated the same as far as the algorithm is concerned. Let the points in the first or experimental data, $S$, be denoted by $\{x_i| i = 1, \ldots, m\}$, and those in the second or model data set, $S'$, be denoted by $\{y_j| j = 1, \ldots, n\}$.

When the registration transformation $T$ is applied to $S'$, the distance to a point on $S$ to $S'$ is zero in the noise-free case. Also, when the inverse of $T$ is applied to $S$, the distance to a point on $S'$ to $S$ should be also zero. The objective of registration is to determine the motion between the two data sets, i.e., $R$ for rotation and $t$ for translation, such that the following criterion is minimized

$$F(R, t) = \frac{1}{\sum_{i=1}^{m} p_i} \sum_{i=1}^{m} p_i d^2(Rx_i + t, S')$$

$$+ \frac{1}{\sum_{j=1}^{n} q_j} \sum_{j=1}^{n} q_j d^2(R^T y_j - R^T t, S')$$

(14)

where $d(x, S)$ denotes the distance of point $x$ to shape $S$ and $p_i$ (respectively $q_j$) takes value one if the point $x_i$ (respectively $y_j$) can be matched to a point on $S'$ in the model data set (respectively $S$ in the experimental data set) and takes value zero otherwise. Again, in the noise-free case, the minimum of Equation (14) will be zero. The relevance and importance of the parameters $p_i$ and $q_j$ are that they enable the algorithm to handle appearance, disappearance, outliers, and occlusion because if a point and its closest point in the other data set are not reasonably paired, then the matching is discarded.

Equation (14) is symmetric in terms of the two data sets, that is, neither data set prevails over the other. One simplification that can be made is to only use the first part of the right hand side of Equation (14). It has been shown in [24] that the effects of the simplification are nominal with respect to the motion estimate but it halves the computation time. The new objective function to be minimized can be stated formally as

$$F(R, t) = \frac{1}{\sum_{i=1}^{m} p_i} \sum_{i=1}^{m} p_i d^2(Rx_i + t, S').$$

(15)

The minimization of Equation (15) is very difficult for two reasons. The first is that $d(Rx_i + t, S')$ is highly nonlinear because the correspondence between $x_i$ and $S'$ is not known beforehand. The second is an Integer Programming Problem, that is, the parameter $p_i$ can take the value zero or one. Zhang [24] used a heuristic approach and assumes the motion between the data sets is small or approximately known. If the motion is approximately known, then the initial motion estimate is applied to the experimental data set and the Zhang algorithm can then be applied to obtain a precise motion estimate. The main problem with the previous technique is its computational complexity. To solve this problem, we present in the following section a genetic algorithm-based method to minimize the motion objective function between the two data sets.

B. Minimization Using Genetic Algorithms

Genetic algorithms (GAs) [30] provide a powerful and domain independent search method for complex, poorly understood search spaces. GAs borrow its name form the natural genetic system. In natural genetic system whether a living cell will perform a specific and useful tasks in a predictable and controlled way is determined by its genetic makeup, i.e., by the instructions contained in a collection of chemical messages called genes [29]. These genes are passed from one generation to the next, so that offspring inherit traits from their parents. According to the "Natural Selection" theory, individuals that are better fit to a given environment are more likely to survive.

GAs use the same tools. Points in the search space are encoded to form strings of a certain elements (1 and 0 for binary coding as an example). These strings in turn form a population and a selection process choose high fitness string according to a certain criterion function. This means that strings with higher fitness value have a higher probability of reproducing new strings in the new generation.

Two operators, crossover and mutation, introduce new individuals (points) into the population (search space). These operations enable GAs to use relatively
few samples to search large spaces. Crossover proceeds in two steps: First chosen strings are mated at random, and then crossover between the two strings occurs at an integer position k along the string. This position is also selected at random. For example if we have two strings $A$ and $B$ such that

$$A = 01101, \quad B = 11000$$

(16)

Let $k = 4$, then the two resulting strings will be

$$A^* = 01101, \quad B^* = 110001.$$  

(17)

Mutation is the occasional (with very small probability) random alteration of the value of a string position. For example $A^*$ after mutation will be

$$A^{**} = 01101.$$  

(18)

In order for the GAS to work we need to define

1. A Fitness function
2. Coding for genes
3. A selection operator
4. A method for crossover and mutation
5. Population of strings
6. A stopping criterion.

To understand how GAS are successful in obtaining the optimal solution, we need to define some terms [29]. The first term is the schemata, $H$, which is a similarity template describing a subset of strings with similarities at certain string positions. In general for string of length $l$ and $k$ alphabet there are $(k + 1)^l$ schemata. The second term is the order of schemata, $O(H)$, which is the number of fixed positions present in the template. Finally the length of a schemata, $d(H)$, which is the distance between the first and last specific string position in the template.

Suppose at time $t$, in a population $A(t)$ we have $m$ examples of a particular schemata, $H$, we call it $m_t(H)$. A string $A_i$ is selected according to a probability, $p_i$, defined as

$$p_i = \frac{f_i}{\sum f_i}. \quad (19)$$

So if a non overlapping population of size $n$ is selected then we expect to have [29]

$$m_{t+1}(H) = m_t(H) n \frac{f(H)}{\sum f_i} \quad (20)$$

where $f(H)$ is the average fitness of $m_t(H)$. Let

$$f_a = \frac{\sum f_i}{n} \quad (21)$$

then

$$m_{t+1}(H) = m_t(H) f(H) \frac{f_a}{f_a} \quad (22)$$

A schemata $H$ is destroyed by crossover with probability $p_d$ defined as

$$p_d = \frac{d(H)}{(l - 1)} \quad (23)$$

or it survives with $p_s = 1 - p_d$. If crossover is random with probability $p_c$, then

$$p_s \geq 1 - p_c \frac{d(H)}{(l - 1)} \quad (24)$$

so after crossover and mutation we expect

$$m_{t+1}(H) > m_t(H) \frac{f(H)}{f_a} (1 - p_c) \frac{d(H)}{(l - 1)} - O(H)p_m \quad (25)$$

where $p_m$ is the mutation probability. Therefore, the probability that a certain schemata will survive is determined by the fitness of the individuals in this schemata. If this fitness is high, which means it contains good solution to the problem, the probability of surviving is high. Also if the length of the schemata is small, the probability of surviving is higher.

Using this fact we implemented this algorithm in order to maximize $F_r(R, t)$, which is the revised $F(R, t)$ described in Equation (26).

$$F_r(R, t) = -\sum_{i=1}^m \sum d^2(Rx_i + t, S'). \quad (26)$$

We used only 5 parameters in $F_r(R, t)$, the three angles of rotation and the displacement in the x and y direction. These parameter were coded in binary. The reason behind using the binary coding was to minimize as possible the length of the schemata $d(H)$ and the order of the schemata $O(H)$. The crossover and mutation operators works separately on each coded parameters with different probabilities, however the generated strings are concatenated together to form one string from which the populations are formed. Results obtained using this registration algorithm are shown in Figure (5).

The GA system for registration is summarized as follows; we use $F_r(R, t)$ as the fitness function, the genes are formed from the concatenation of the five binary coded parameters, the three angles of rotation that forms the matrix $R$ and the x an y translations parameter. The selection operator chooses the highest fitted genes for mating using a Roulette wheel
The main advantages of this system is that the equipments used are already available in most dentist offices. The color images could be further processed to obtain a texture mapping of the jaw model which will enable detection and analysis of gum diseases.

This technique was able to register the data points of two consecutive shots in less than 3 minutes compared to 24 minutes using the Zhang algorithm. Table I presents a comparison between GAs performance and other registration techniques. [26]

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Execution Time minutes:seconds</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang</td>
<td>32:36</td>
<td>0.1995</td>
</tr>
<tr>
<td>Zhang/k-D tree</td>
<td>24:50</td>
<td>0.2000</td>
</tr>
<tr>
<td>ICP/k-D tree</td>
<td>36:23</td>
<td>0.1989</td>
</tr>
<tr>
<td>GAs</td>
<td>2:55</td>
<td>0.2000</td>
</tr>
</tbody>
</table>

V. CONCLUSIONS

In this paper, we presented a novel approach is to obtain a record of the patient’s occlusion using computer vision. Data acquisition was conducted using visible light and intra-oral video cameras. The technique utilizes shape from shading to record and obtain the "electronic study models" which will then be available for diagnostic, treatment planning, and treatment purposes. The contribution of this paper is part of a model-based vision system for orthodontics that is designed to replace plaster models.

The main advantages of this system is that the equipments used are already available in most dentist offices. Using a calibration procedure accurate measurements from the reconstructed 3D model could be obtained. The data acquisition takes only few minutes which will reduce effectively the inconvenience for both the patient and the orthodontist. The color images could be further processed to obtain a texture mapping of the jaw model which will enable detection and analysis of gum diseases.

The 3D model of the jaw could be further processed to obtain each tooth separately, determine its position, presence/absence and orientation. Together with the radiography of the jaw an accurate mechanical model of the tooth and its root could be determined and we this could be used to predict the kinetic and kinematics of the tooth in tooth corrections.

REFERENCES

[16] Ping-Sing,Tsai, and Mubarak. Shape from shading using linear approximation. Tech. report, University of Center Florida.


Fig. 4. Extracting 3D points from a sequence of 2D images using SFS

Fig. 5. Wireframe and rendered surface of the registration results

Fig. 6. Part of the Jaw model obtained from 15 data sets