

# A Robust 3-D Reconstruction System for Human Jaw Modeling

Sameh M. Yamany<sup>1</sup>, Aly A. Farag<sup>1</sup>  
David Tasman<sup>2</sup>, and Allan G. Farman<sup>2</sup>

<sup>1</sup> Computer Vision and Image Processing Laboratory, University of Louisville

<sup>2</sup> School of Dentistry, University of Louisville, Louisville, KY 40292, USA  
farag@cvip.uofl.edu

**Abstract.** This paper presents a model-based vision system for dentistry that will assist in diagnosis, treatment planning and surgical simulation. Dentistry requires the accurate 3-D representation of the teeth and jaws for diagnostic and treatment purposes. The proposed integrated computer vision system reconstructs a 3-D model of the patient's dental occlusion using an intra-oral video camera. A modified shape from shading (SFS) technique, using perspective projection and camera calibration, extracts the 3-D information from a sequence of 2-D images of the jaw. Data fusion and 3-D registration techniques develop the complete jaw model. Triangulization is then performed, and a solid 3-D model is obtained via rapid prototyping. The system performance is investigated using ground truth data, and the results show sub-millimeter reconstruction accuracy. The system is shown to be robust in terms of speed and accuracy compared to current practices.

## 1 Introduction

Dentistry requires the accurate 3-D representation of the teeth and jaws for diagnostic and treatment purposes. For example, orthodontic treatment involves the application, over time, of force systems to teeth to correct malocclusion. In order to evaluate tooth movement progress, the orthodontist monitors this movement with visual inspection, intra-oral measurements, fabrication of plastic models (casts), photographs, and radiographs; this process is both costly and time consuming. Moreover, repeated acquisition of radiographs may result in untoward effects. Obtaining a cast of the jaw is a complex operation for the dentist, an unpleasant experience for the patient, and may not provide all the necessary details of the jaw.

Oral and maxillofacial radiology can provide the dentist with abundant 3-D information of the jaw. Current and evolving methods include computed tomography (CT) [1], tomosynthesis, tuned-aperture CT (TACT), and localized, or "cone-beam," computed tomography [2]. While oral and maxillofacial radiology is now widely accepted as a routine technique for dental examinations, the equipment is rather expensive and the resolution is frequently too low for

3-D modeling of dental structures. Furthermore, the radiation dose required to enhance both contrast and spatial resolution can be unacceptably high.

Recently, efforts have focused on computerized diagnosis in dentistry [3]. Usually, most of the 3-D systems for dental applications found in the literature rely on first obtaining an intermediate solid model of the jaw (cast or teeth imprints) and then capturing the 3-D information from that model. User interaction is needed in such systems to determine the 3-D coordinates of fiducial reference points on a dental cast. Other systems measuring the 3-D coordinates have been developed using either mechanical contact or a traveling light principle [4]. Generally, such systems are either not accurate or are time and labor intensive.

The authors have been involved for the last five years in a project to develop a system for dentistry to go beyond traditional approaches in diagnosis, treatment planning, surgical simulation, and prosthetic replacements. Specific objectives are as follows: (i) to design a data acquisition system that can obtain sequences of calibrated video images of the upper/lower jaw using small intra-oral cameras with respect to a common reference in 3-D space; (ii) to develop methods for accurate 3-D reconstruction from the acquired sequence of intra-oral images. This involves using a new algorithm for shape from shading (SFS) that incorporates the camera parameters; (iii) to develop a robust algorithm for the fusion of data acquired from multiple views, including the implementation of an accurate and fast 3-D data registration; (iv) to develop a specific object segmentation and recognition system to separate and to recognize individual 3-D tooth information for further analysis and simulations; and (v) to develop algorithms to study and to simulate tooth movement based on the finite element method and deformable model approaches. This research will have an immense value in various dental practices including implants, tooth alignment, and craniofacial surgery. The research will also have wide applications in dental education and training.

This paper describes the project's first phase concerning the development of a 3-D model of the jaw, not from a cast, but from the actual human jaw. The work reported here is original and novel in the following aspects: (1) data acquisition is performed directly on the human jaw using a small off-the-shelf solid state camera, (2) the acquisition time is relatively short and is less discomforting to the patient compared to current practices, (3) the acquired digital model can be stored with the patient data and retrieved on demand, (4) these models can also be transmitted over a communication network to remote sites for further assistance in diagnosis and treatment planning, and (5) dental measurements and virtual restoration can be performed and analyzed. This work also involves three important areas in the computer vision and medical imaging fields, namely: shape recovery, data fusion, and surface registration.

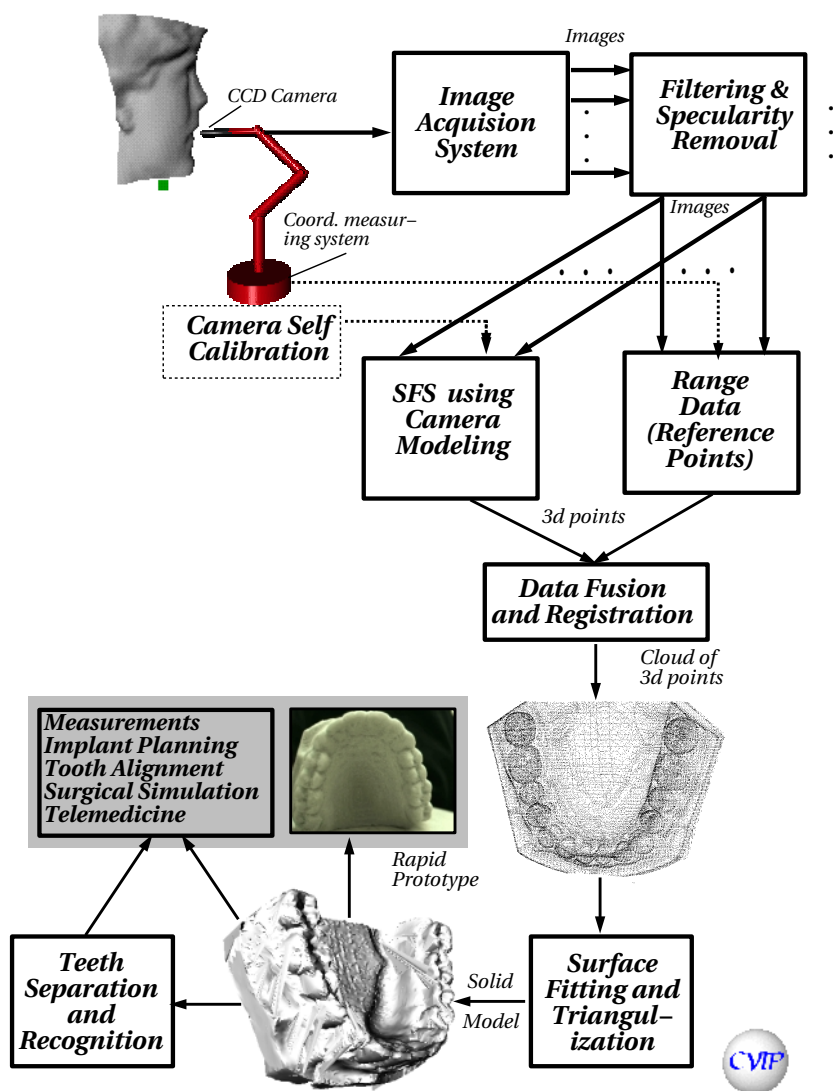
## 2 System Overview

As shown in Fig 1, our approach to reconstruct the human jaw consists of the following stages. The first stage is data acquisition. A small intra-oral Acu-

Cam(Dentsply/New Image, Canaga Park, California) CCD camera with built-in laser light, is calibrated and then placed inside the oral cavity. The camera acquires a set of overlapping images  $\{I_j \mid j = 1, 2, \dots, J\}$  for different parts of the jaw such that  $\bigcup_{j=1}^J I_j$  covers the whole jaw. The images are preprocessed to reduce noise, sharpen edges, and remove specularities. Specularity removal was done using an approach similar to the Tong and Funt technique [5].  $J$  sets of points are then computed with a modified SFS algorithm, which accounts for the camera perspective projection. To obtain accurate metric measurements, range data is obtained using a five link digitizer arm. These data consist of some reference points on the jaw. Fusion of the range data and the SFS output provides accurate metric information that can be used later for orthodontic measurements and implant planning. A fast registration technique is required to merge the resulting 3-D points to obtain a complete 3-D description of the jaw [6]. The final stage is to transform this model into patches of free form surfaces using a triangulization technique. This step enables the development of a 3-D solid model via rapid prototyping. Further processing on the digital model includes tooth separation, force analysis, implant planning, and surgical simulation.

### 3 Shape from Shading Using Perspective Projection and Camera Calibration

Among the tools used in shape extraction from single view is the shape from shading (SFS) technique. The surface orientation at a point  $\mathbf{M}$  on a surface  $S$  is determined by the unit vector perpendicular to the plane tangent to  $S$  at  $\mathbf{M}$ . Most of the research done in SFS assumes orthographic projection from which the elemental change in the depth  $Z$  at an image point  $(x, y)$  can be expressed as  $\delta z \approx \frac{\partial Z}{\partial x} \delta x + \frac{\partial Z}{\partial y} \delta y$ . The partial derivatives are called surface gradients  $(p, q)$ . The surface normal to a surface patch is related to the gradient by  $\mathbf{n} = (p, q, 1)$ . 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:  $E(x, y) = R(p, q, \mathbf{L})$ , where  $R(p, q, \mathbf{L})$  is the surface reflectance map and  $\mathbf{L}$  is the illuminant direction. Many algorithms were developed to estimate the illuminant direction [7]. Because the laser light beam is built in the CCD camera, we assume that this will be the only source of light inside the mouth cavity and that the illuminant direction is known beforehand. However, the assumption of orthographic projection is not adequate as the camera is very close to the object. We propose to calibrate the CCD camera and use the perspective projection matrix to enhance the SFS algorithm and to obtain a metric representation of the teeth and gum surfaces. The perspective projection equation is as follows:



**Fig. 1.** The process starts by capturing a sequence of video images using a small intra-Oral CCD camera. These images are preprocessed to remove specularity. Reference points are obtained using the CMM system. SFS is applied to the images. The range data are fused to the SFS output and then registration takes place. A cloud of points representing the jaw is obtained and, by triangulization, a solid digital model is formed. This model is reproduced using a rapid prototype machine. Further analysis and orthodontics application can be performed on the digital model.

$$s\mathbf{m} = \mathbf{B}\mathbf{M} + \mathbf{b} \quad \text{or,} \quad (1)$$

$$\mathbf{M} = \mathbf{B}^{-1}(s\mathbf{m} - \mathbf{b}) = f(s(x, y)) \quad (2)$$

where  $\mathbf{B}$  is a  $3 \times 3$  matrix and  $\mathbf{b}$  is a translation vector. The matrix  $[\mathbf{B}\mathbf{b}]$  is called the perspective projection matrix. The function  $f(s(x, y))$  maps  $\mathbf{M}$  to a point  $\mathbf{m}$  in the image. The normal to the surface at  $\mathbf{M}$  is defined to be the cross product of the two gradient vectors  $\mathbf{p} = \frac{df(s(x, y))}{dx}$ ,  $\mathbf{q} = \frac{df(s(x, y))}{dy}$ . The surface reflectance  $R(\cdot)$  becomes a function of the scalar  $s$  defined in equation[1] as follows,

$$R(s) = \frac{(\mathbf{p} \times \mathbf{q}) \cdot \mathbf{L}}{|\mathbf{p} \times \mathbf{q}| |\mathbf{L}|} \quad (3)$$

The new formulation of the SFS problem becomes finding the scalar  $s$  that solves the new brightness equation  $g(s) = E(x, y) - R(s) = 0$ . This can be solved using a Taylors series expansion and applying the Jacoby iterative method [8] where at the  $n^{th}$  iteration, for each point  $(x, y)$  in the image,  $s_{x,y}^n$  is as follows:

$$s_{x,y}^n = s_{x,y}^{n-1} + \frac{-g(s_{x,y}^{n-1})}{\frac{d}{ds_{x,y}}g(s_{x,y}^{n-1})} \quad \text{where,} \quad (4)$$

$$\frac{d}{ds_{x,y}}g(s_{x,y}^{n-1}) = -\frac{d\mathbf{N}}{ds_{x,y}} \cdot \frac{\mathbf{L}}{|\mathbf{L}|} \quad (5)$$

$$\frac{d\mathbf{N}}{ds_{x,y}} = \frac{d\mathbf{v}}{ds_{x,y}} \frac{1}{\sqrt{\mathbf{v}^t \mathbf{v}}} - \frac{\mathbf{v}}{\sqrt{(\mathbf{v}^t \mathbf{v})^3}} \left( \mathbf{v}^t \frac{d\mathbf{v}}{ds_{x,y}} \right) \quad (6)$$

$$\frac{d\mathbf{v}}{ds_{x,y}} = \mathbf{B}^{-1}\mathbf{m} \times \mathbf{B}^{-1}(0, s_{x,y-1}, 0)^t + \mathbf{B}^{-1}(s_{x-1,y}, 0, 0)^t \times \mathbf{B}^{-1}\mathbf{m} \quad (7)$$

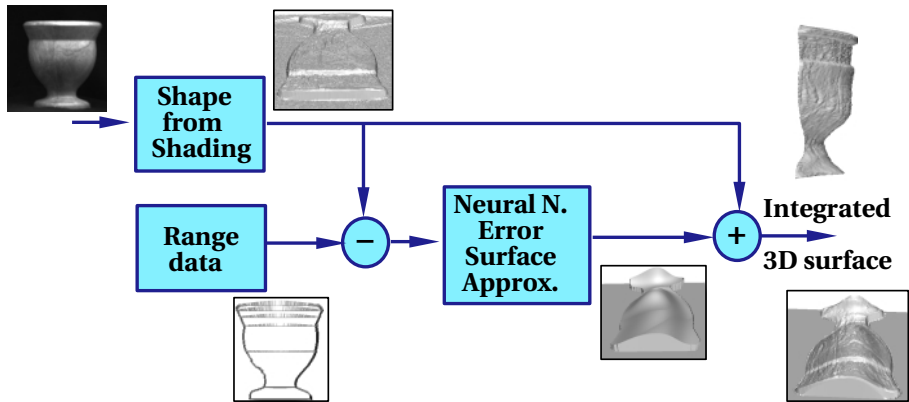
where  $\mathbf{v} = \mathbf{p} \times \mathbf{q}$ .

Even though camera parameters are used in the SFS implementation, accurate metric information cannot be deduced from the resulting shape because only one image is used. Additional information is needed to complement the SFS output and to incorporate the metric measurements.

## 4 Fusion of SFS and Range Data

The most important information for reconstructing an accurate 3-D visible surface, missing in SFS, is the metric measurement. SFS also suffers from the discontinuities due to highly textured surfaces and different albedo. The integration of the dense depth map obtained from the SFS with sparse depth measurements obtained from a coordinate measurement machine (CMM) for the reconstruction of 3-D surfaces with accurate metric measurements has two advantages [9]. First, it helps in removing the ambiguity of the 3-D visible surface discontinuities produced by shape from shading. Second, it complements the missing metric information in the SFS. The integration process, as depicted in Fig 2, includes

the following. First, the error difference in the available depth measurements between the two sensory data are calculated. Next, a surface that fits this error difference is approximated. Finally, the approximated surface is used to correct the SFS.



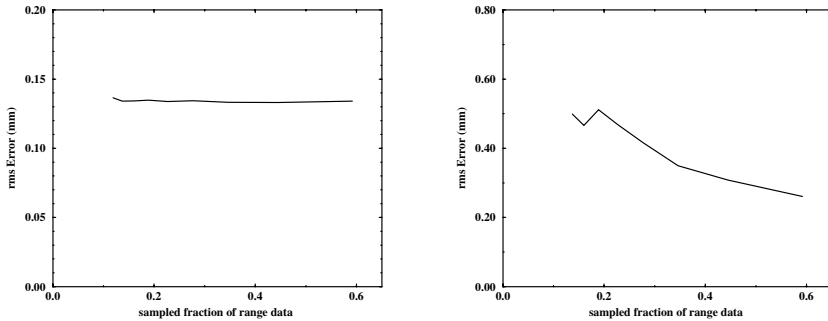
**Fig. 2.** Functional block diagram of the integration process of SFS and range data. The surface approximation process uses neural networks. An example of using this system on the image of a vase is demonstrated.

We used a multi-layer network for the surface approximation. The  $x$ - and  $y$ -coordinates of the data points are inputs to the network, while the error in the depth value at the point  $(x, y)$  is the desired response. The learning algorithm used is the error Kalman-filter learning technique because of its fast weights computation. The error difference between the SFS and the range measurements and their  $x$ - $y$  coordinates are used to form the training set. The input to the network is the  $x$ - $y$  coordinates and the output is the error difference at that coordinate. After training, the neural network provides the approximated smooth surface that contains information about the errors in the SFS at the locations with no range data. This approximated surface is then added to the SFS. The result is the 3-D surface reconstruction containing accurate metric information about the visible surface of the sensed 3-D object. An example performed on the image of a vase is shown in Fig 2.

The output of the fusion algorithm to each image is a set of 3-D points describing the teeth surface in this segment. However, there is no relation between the 3-D points of a segment and the following segment. Thus we needed a fast and accurate 3-D registration technique to link the 3-D points of all the segments to produce one set of 3-D points describing the whole jaw surface. Yamany et. al. [6,10] introduced a new 3-D registration technique using the Grid Closest Point (GCP) transform and Genetic Algorithms (GA). This technique was faster and more accurate than the existing techniques found in the literature.

## 5 Validation

To validate the SFS and data fusion methods and to determine the accuracy of the proposed system and the required resolution, we used a ground truth dense depth map registered with intensity images obtained from a laser range scanner. Although laser scanners have accuracy limitation, their output can be considered as ground truth since they are now widely accepted in the dentistry practices. The RMS error between the integrated surface and the ground truth is used as a measure of the system performance. Two different types of surfaces, a sphere as a smooth surface and a free-form surface (see Mostafa et al. [9] for more details), are investigated. Figure 3 shows the results of this analysis. The RMS error for the smooth surface is smaller compared to that of the free-form surface. This is an expected result because the surface approximation process tends to smooth the surface where range data are not available, producing a large RMS error in the case of free-form surfaces. The results show that a higher sampling rate increases the accuracy in the case of free-form surfaces, yet this will increase the time to acquire the data. However, the sampling rate has minimal effect in the case of smooth surfaces. The results and the above analysis show that the system can achieve sub-millimeter accuracy with a small number of reference points.



**Fig. 3.** The RMS error between the integrated surface and the ground truth is used as a measure of the system performance. Two different types of surfaces:(left) a sphere as a smooth surface and (right) a free-form surface are investigated.

## 6 Results and Discussions

Intra-Oral cameras are quickly becoming standard equipment used by many dental practitioners [11]. In our experiments, we used an intra-oral CCD cam-

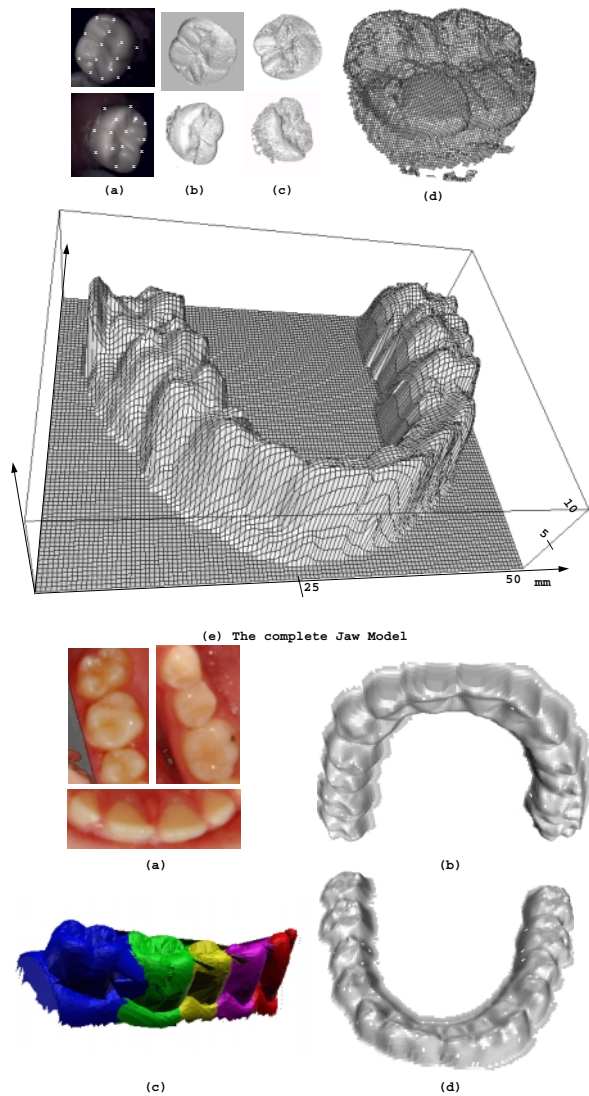
era, AcuCam (Dentsply/ New Image Inc.). The experiments were conducted on different pediatric subjects at the Orthodontics Department, University of Louisville, KY. After calibrating the camera, a sequence of images capturing segments of the jaw are obtained. The process of taking the images was relatively fast, taking about 4 to 5 minutes to cover the upper/lower jaw. The images were taken carefully to cover all visible surfaces of the teeth. With the CMM system (with a resolution of 0.23 mm, and sampling rate of 1,000 points/second), reference points were calculated for each image. Figure (a) shows an example of two images taken of a patient's tooth. The complete tooth surface is covered in these two images. Figure (b) shows the outputs of the SFS stage. Using the range data shown as cross signs in Fig (a) and applying the fusion algorithm results in the corrected surfaces shown in Fig (c). The registration procedure on these two data sets provides the complete surface model of the tooth as shown in Fig (d). A smooth version of the whole jaw model is shown in Fig (e). This model contains all required metric information, and can be used to measure any orthodontic parameter and can be reproduced with the same original scale. More results of applying the reconstruction algorithms on another subject are shown in Fig (f,g,h). The resulting jaw models have sub-millimeter accuracy and are faithful enough to show all the information about the patient's actual jaw in a metric space. Both the time and convenience for the patient must be considered when comparing these results with those from scanning a cast. Further processing was performed on the digital jaw model, and Fig (i) shows the result of the teeth segmentation and identification stage.

## 7 Conclusions and Future Extensions

The 3-D reconstruction of the human jaw has tremendous applications. The model can be stored along with the patient data and can be retrieved on demand. The model can also be used in telemedicine where it is transmitted over a communication network to different remote dentists for further assistance in diagnosis. Dental measurements and virtual restoration could be performed and analyzed. This work enables many orthodontics and dental imaging research projects, applied directly to the jaw and not to a cast, using computer vision and medical imaging tools. The paper describes the results of the first phase in a project aimed to enhance current dental imaging practices.

The next phase includes the analysis and simulation of dental operations including tooth alignment, implant planing, restoration, and measurement of distances and orientation of teeth with respect to each other. Similar work was performed by Alcaniz et al. [12]. However, their analysis was done on the tooth contours and not on the actual 3-D model of the tooth. Also, the arch-wire model does not account for the 3-D displacements of the tooth.





**Fig. 4.** (a) Intra-oral intensity images from one patient with range data marked as cross signs. (b) 3-D visible surfaces obtained from the SFS. (c) The final surfaces obtained from the integration process. (d) A visible surface mesh obtained from registering the two views in (c). (e) A smoothed version of the whole jaw model. (f) Some intensity images taken from another patient. (g-h) The upper and lower digital jaw models of this patient. (i) The result of the teeth segmentation and identification stage.

## Acknowledgment

This work was supported in part by grants from the Whitaker Foundation, and the NSF (ECS-9505674) institutions.

## References

1. P. van der Stelt and S. M. Dunn, "3d-imaging in dental radiography," in *Advances in Maxillofacial Imaging*, A. G. Farman, ed., pp. 367–372, Elsevier Science B. V., 1997. 778
2. R. L. Webber, R. A. Horton, D. A. Tyndall, and J. B. Ludlow, "Tuned-aperture computed tomography (tact). theory and application for three-dimensional; dento-alveolar imaging," *Dentomaxillofac. Radiol.* **26**, pp. 51–62, 1997. 778
3. D. Laurendeau and D. Possart, "A computer-vision technique for the aquisition and processing of 3-d profiles of dental imprints: An application in orthodontics," *IEEE Transactions on Medical Imaging* **10**, pp. 453–461, Sep 1991. 779
4. A. A. Goshtasby, S. Nambala, W. G. deRijk, and S. D. Campbell, "A system for digital reconstruction of gypsum dental casts," *IEEE transactions on Medical Imaging* **16**, pp. 664–674, Oct 1997. 779
5. F. Tong and B. V. Funt, "Removing speculariry from color images for shape from shading," in *Computer Vision and Shape Recognition*, A. Krzyzak, T. Kasvand, and C. Y. Suen, eds., vol. 14 of *Computer Science*, pp. 275–290, World Scientific, 1989. 780
6. S. M. Yamany, M. N. Ahmed, and A. A. Farag, "A new genetic-based technique for matching 3d curves and surfaces," *Pattern Recognition (to appear)*, 1999. 780, 783
7. A. Pentland, *Extract Shape From Shading*, Academic Press, MIT Media Lab, 2nd ed., 1988. 780
8. P. S. Tsai and M. Shah, "A fast linear shape from shading," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 734–736, July 1992. 782
9. M. G.-H. Mostafa, S. M. Yamany, and A. A. Farag, "Integrating shape from shading and range data using neural networks," *Proc. IEEE Int. Conf. Comp. Visi. Patt. Recog. (CVPR)*, June 1999. Fort Collins, Colorado. 782, 784
10. S. M. Yamany, M. N. Ahmed, and A. A. Farag, "Novel surface registration using the grid closest point (gcp) transform," *Proc. IEEE International Confenrence on Image Processing, Chicago* **3**, pp. 809–813, October 1998. 783
11. L. A. Johnson, "A systematic evaluation of intraoral cameras," *Journal of the California Dental Association* **22**, pp. 34–47, November 1994. 784
12. M. Alcaniz, C. Montserrat, V. Grau, F. Chinesta, A. Ramon, and S. Albalat, "An advanced system for the simulation and planning of orthodontic treatment," *Medical Image Analysis* **2**, pp. 37–60, March 1998. 785