COMPUTERIZED APPROACHES TO THE OBJECTIVE ASSESSMENT OF CRANIOFACIAL SHAPE

John Henry DiLiberti
A.B., Princeton University, 1966
M.D., University of Cincinnati College of Medicine, 1970

A thesis presented to the faculty of the
Oregon Graduate Institute of Science & Technology
in partial fulfillment of the
requirements for the degree
Master of Science
in
Computer Science and Engineering

July 1993

The thesis "Computerized Approaches to the Objective Assessment of Craniofacial Shape" by John H DiLiberti has been examined and approved by the following Examination Committee:

STEVEN OTTO

Assistant Professor

Thesis Advisor

RONALD COLE

Professor

JONATHAN ZONANA

Professor

Oregon Health Sciences University

ACKNOWLEDGMENTS

The support and encouragement of Lise Storc and David Olson, Ph.D. are gratefully acknowledged. Their constructive criticism was very helpful to me during the development of a number of the ideas contained in this paper. I wish Dave the best of luck as he moves from a career in theoretical physics to medicine.

DEDICATION

My family has always been a source of strength, particularly my children John and Lara, and my parents who must wonder if I'll ever finish school. I am especially appreciative of my wife Cindy, without whose support I may never have finished this project.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Computerized Approaches to the Objective Assessment of Craniofacial Shape

John H. DiLiberti, M.D.

Human facial appearance plays an important role in medical diagnosis. In medical genetics, pain management, and psychiatry, judgments about the configuration of the face may substantially influence clinical evaluation and treatment. Often these judgments are entirely subjective, based on the observer's experience, skill, and intuition - leading to considerable diagnostic variability. Recent interest in improving objectivity has been driven in part by advances in molecular genetics. Successful application of molecular technology is critically dependent upon accurate clinical diagnosis. Classical anthropometric techniques and related photogrammetry have been used with some success to evaluate facial configuration objectively, but they provide at best limited data about a highly complex surface.

The computational approach to evaluation of facial configuration will be dependent upon the choice of a 2-dimensional or 3dimensional data model. Two-dimensional images are much easier to obtain, but obviously entail the loss of information. Preliminary investigation showed that neural networks could be trained to differentiate between smiling and sad faces. They may therefore be useful in behavioral areas such as the evaluation of pain in non-verbal patients. Medical genetics, however, probably requires three-dimensional information. A photograph may not even suggest a diagnosis which is easy to make while examining the patient. Three-dimensional contours of the face have been obtained by MRI and CT scanning along with stereophotometric techniques. Problems associated with these approaches include radiation exposure, requirement for sedation to maintain constant position, expense, and inability to use the required equipment in many clinical settings. In addition, the face of an infant or young child is so lacking in topographic features that computerized matching of corresponding pixels in the left and right images is almost impossible at any usable degree of resolution.

Accordingly an imaging technique using projected structured light and a single camera was developed. Parallel bands of white light had been used previously with some degree of success, but areas of the face with low reflectivity such as eyelashes and eyebrows made it difficult to impossible to trace dark bands in those regions. Preliminary results showed that projecting band patterns with multiple colors could drastically reduce this problem. Remaining technical problems included creation of projection filters with sufficient spectral purity and the differential sensitivity of the three color layers in photographic film. Careful selection of projected colors led to successful extraction of the 3D coordinates from the deformation of the band pattern by the facial surface. With improved filters and imaging capabilities it could be possible to depth-encode the structured light spectrally so that the z coordinate could be determined directly from the color of the pixel. After three-dimensional surface contours of the face become available, methods to analyze this information must be developed. Successful application of this approach could lead to important advances in the diagnosis of genetic disorders.

CHAPTER 1

INTRODUCTION

The face is an important and complicated component of human anatomy. Although the major structural modifications which transform embryonic tissues into a fully-differentiated, but immature, face are essentially complete by the 14th week of a 40 week gestation [Moore 1973] (Figure 1.1),

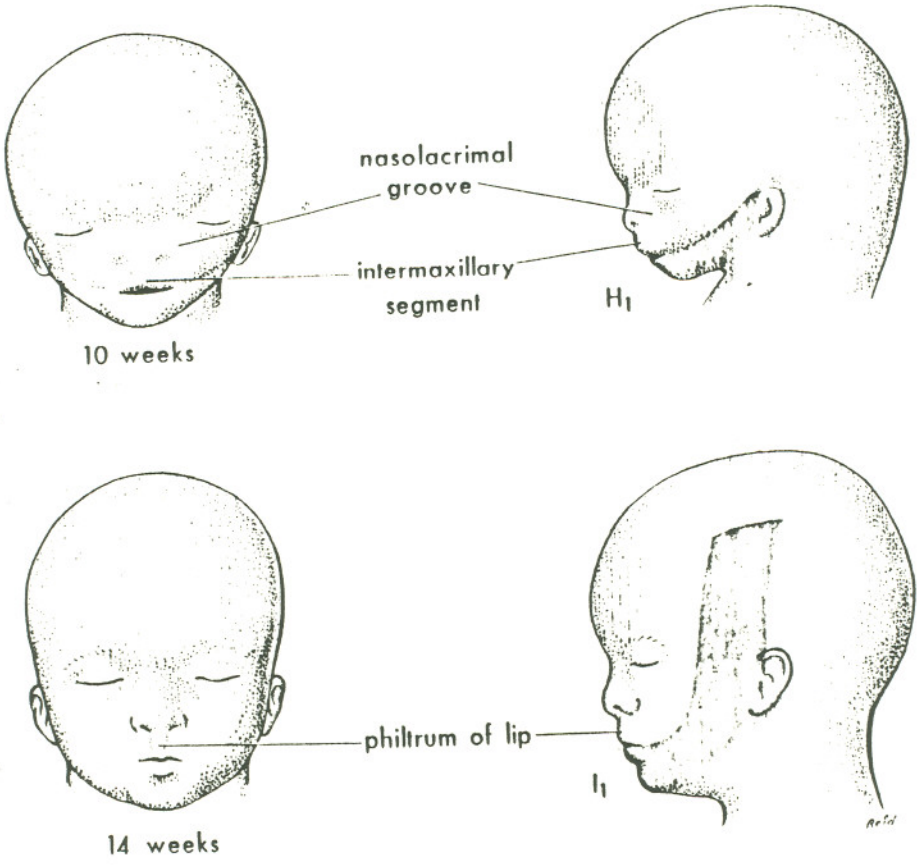


Figure 1.1 The embryology of the human face at 10 and 14 weeks. [Moore]

The resting surface configuration (i.e. with facial musculature in a neutral state with typical degree of tone) continues to change throughout the remainder of the pregnancy and, indeed, for the duration of life. The forces which determine final fetal facial configuration are as yet poorly understood, but are clearly driven by genetic factors and may at times be affected by external pharmacologic, microbial, and mechanical agents.

In addition to the obviously wide social and cultural importance of human facial configuration, biomedical implications also abound. The resting configuration is of considerable importance both in the diagnosis of birth defects, and also in the surgical correction of congenital and acquired abnormalities, while changes in configuration as a consequence of expressive behaviors appear to be very significant in fields such as psychiatry, psychology, and the diagnosis and treatment of pain.

Although a substantial amount of work has gone into the development of methods dealing with certain selected aspects of craniofacial configuration, the major diagnostic and therapeutic advances in health care which have become available during the past one or two decades make it very likely that improved techniques for objective craniofacial assessment will become either highly desirable or even necessary, in order to apply these advances most effectively. This is particularly true of the application of molecular genetics* to disorders which include morphologic abnormalities of the craniofacial region.

It should be noted that this discussion is concerned primarily with the surface configuration of the face. Although the surface is, to a significant extent, dependent upon the relationships of the soft tissues to the underlying bony structures, the clinical areas of interest under consideration here - predominantly medical genetics with some applications for the behavioral sciences - are, for a number of reasons, focused on the visual appearance of the face. The underlying 3 dimensional structures may eventually turn out to be far more important diagnostically, but for the time being are less accessible (since they require "invasive" procedures to image) and clinicians do not find examining these images as intuitive as looking directly at a face.

---

* (Molecular genetics is used here as an all-encompassing term to describe the study of genes using technology which examines the actual DNA, either directly or by inference. Biochemical genetics, on the other hand, generally looks at the metabolic processes regulated by the genes.)

Eventually it may be feasible to obtain images of a face and either make a diagnosis directly using computerized analysis or, more likely, use the computerized analysis to give the clinician a more objective assessment of the facial configuration and by comparing to similar objective assessments of the face for known disorders aid in the diagnostic process. The experimental approaches described below are intended to be some initial steps in this direction and reflect some of the challenges discovered along the way. In particular, the difficulty of obtaining accurate surface coordinates under clinically acceptable circumstances was not appreciated early on and led to that research being the bulk of what was accomplished. Actually analysis of the surfaces will be a subsequent challenge.

CHAPTER 2

BIOMEDICAL BACKGROUND AND MOTIVATIONS

Between 2 and 5 % of all children are born with a significant physical abnormality or birth defect. [Jones 1988] Some of these birth defects are isolated anomalies, such as a failure of the tissue precursors to the lips to fuse in the midline - resulting in a cleft lip (formerly referred to colloquially as a "hare lip"). Others are more global and involve multiple structures with anatomical relationships of varying complexity. Among this latter group of disorders, those involving the face and head (commonly called craniofacial syndromes) are of particular interest.[Gorlin et al. 1990] A large number, probably as many as a thousand or more distinct conditions, have been described and virtually every month a variety of medical journals contain numerous additional reports.

For certain syndromes with craniofacial manifestations the etiology is well-characterized, for example in trisomy 21 or the Down syndrome[#] the affected infant has unfortunately received 3 copies of chromosome[@] number 21 instead of the usual 2. The resulting imbalance in genetic regulation produces, in addition to mental retardation and a variety of structural defects involving important organs and the extremities, a characteristic facial appearance, often recognizable in the newborn period, and virtually impossible to miss as the child becomes older. In other clinical situations a syndrome may result from an abnormality of a single gene, rather than all or part of a chromosome as in trisomy 21 and related chromosomal conditions. Still other craniofacial syndromes result from fetal

---

[#] (This condition was previously referred to by a term which has fortunately fallen into disfavor - "mongolism". [Jones 1988] The possessive form Down's is no longer considered to be semantically correct)

[@] (A chromosome (literally "colored body" for their appearance using certain stains under a light microscope) is a complex structure found in the nucleus of cells. It consists of assorted proteins along with an entire package of genes encoded in two long strands of DNA. Human cells have a modal number of 46 chromosomes, one each for chromosome numbers 1 through 22 and two X's in females and an X and Y in males.)

exposure to noxious organisms or chemicals which are transferred from the mother - infections, prescription and illicit drugs, and occupationally or environmentally derived exposures. Finally, some infants apparently suffer from problems which are postulated to result from local, presumably random, sporadic derangements in the fetus, surrounding membranes, and/or uterus, or from disorders which are so poorly understood that no plausible etiology has been suggested.

As a consequence of the heterogeneous nature of the etiologies of craniofacial syndromes the clinical geneticist (sometimes called a dysmorphologist) who is interested in the diagnosis, treatment, and/or research into the pathophysiology and etiology of syndromes affecting craniofacial shape or morphogenesis often struggles with a difficult task. While tests for microscopically visible chromosomal anomalies are readily accessible, and molecular approaches may now be used to identify submicroscopic chromosomal abnormalities as well as single gene defects, confirmatory laboratory tests are presently unavailable for the majority of the known disorders of craniofacial morphogenesis. The dysmorphologist is frequently called on, therefore, to decide upon a clinical diagnosis, a treatment plan, and advise the family about recurrence risks for the syndrome in future offspring. The diagnosis is generally based on visual inspection of the face and head, supplemented by examination of the remaining anatomy of the patient, relevant laboratory tests, and, perhaps, a few point-to-point measurements such as the distance between the eyes, which attempt to capture some of the shape information.

Although human beings are quite good at visual pattern recognition, the complexity of human facial structures makes accurate diagnosis a daunting task. Skilled dysmorphologists not infrequently disagree about whether two individuals have similar or dissimilar facial appearances. Obviously in situations where the prognosis would vary greatly depending upon which of several craniofacial configuration-based diagnoses is present, accuracy becomes extremely important. Families might alter treatment decisions for an affected child depending on the expected long-term outcome, their reproductive decisions might be significantly influenced by the predicted recurrence risks for the condition, or they might make employment or life-style changes to cope with expected needs.

Despite the enormous advances in the understanding of molecular genetics which have been made in the past few years, the molecular approach is not a panacea. Indeed, in the absence of good clinical data, it is virtually impossible to get good (or frequently even

any) molecular insights into a condition. Thus, virtually any approach which may ultimately lead to prevention or better treatment of these conditions is also dependent upon accurate diagnosis. Attempts to map and identify a causative gene will be difficult, or more likely impossible, if studies are done on children who have somewhat similar appearances but, in fact, have conditions with different underlying etiologies. The dysmorphologist is very much in need of tools, therefore, which can be brought to bear diagnostically on craniofacial appearance in an objective fashion.

In addition to structural craniofacial abnormalities which result from genetic and/or environmental factors and those acquired from the destructive effects of either intrinsic disease processes or extrinsic trauma, the surface configuration of the face changes in response to underlying emotional states. These changes are of particular interest in the study of pain in children and also the study of psychiatric disorders in children and adults. The study of pain in early childhood is particularly hampered by children's inability to verbalize how they feel. Substantial evidence shows, however, that facial configuration shows a strong correlation with the experience of pain - even in very young, almost newborn infants. [Grunau & Craig 1987; Young & Ellis 1989] (Figure 2.1)

Figure 2.1 Facial configurations of newborn infants.[Grunau & Craig 1987]

CHAPTER 3

APPROACHES TO THE OBJECTIVE ASSESSMENT OF CRANIOFACIAL SHAPE

As noted above, dysmorphologists occasionally borrow the tools of the physical anthropologist and take various measurements, typically point-to-point dimensions, but various angles and perimeters or circumferences are also used. (Figure 3.1)

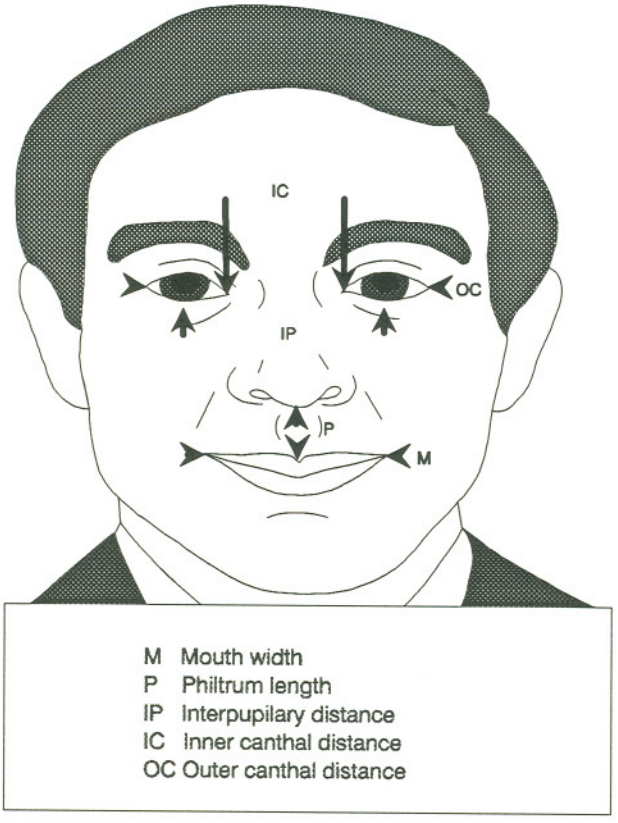


Figure 3.1 Schematic of anthropometric measurements

Although these anthropometric approaches probably add some objectivity to dysmorphologic diagnoses, they have, in fact, not been well standardized. Despite published charts or graphs of norms based on age, few, if any, studies have looked at variations in measurements made by the same observer on different occasions or inter-observer variations and few efforts have been made to correlate anthropometric measurements with head size and body size. [Hall et al. 1989; Pober et al. 1993]

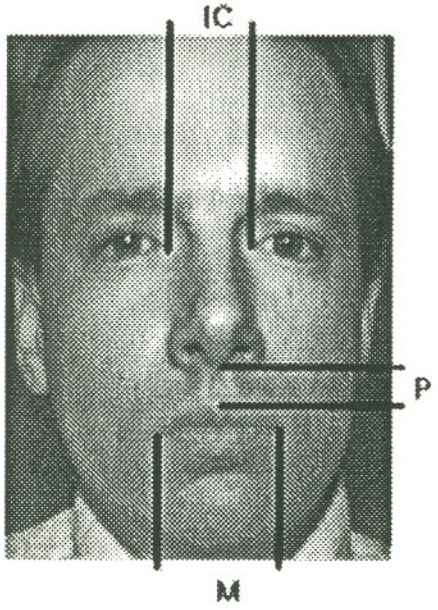
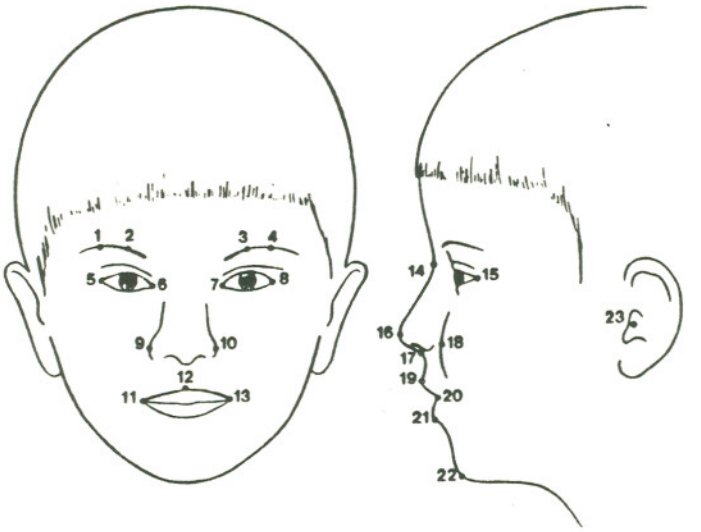


Figure 3.2 Examples of photogrammetric measurements

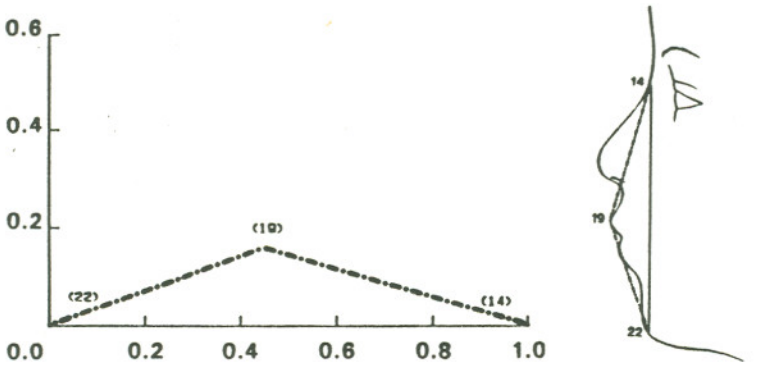
A somewhat similar approach, photogrammetry, may also be used to obtain objective measurements of craniofacial features, but suffers from some of the same limitations. This technique depends upon taking photographs under carefully controlled conditions and then making measurements on the photographs rather than the patient. [DiLiberti & Olson 1991; DiLiberti et al. 1993] (Figure 3.2)

Obvious advantages include the permanent record of the face and the lack of a moving target - particularly helpful with small children. For measurements dependent upon palpation of underlying bony structures, however, photogrammetry may have some shortcomings. Although several theoretical objections have been raised and a few attempts at correlation between the 2 approaches, photogrammetry has never been rigorously compared to anthropometry. [DiLiberti et al. 1993]

Despite the criticisms regarding anthropometry and photogrammetry noted above, they have been shown to have some value for improving objectivity in clinical diagnosis. Bookstein [Bookstein 1978; Bookstein 1983] has described mathematically and Clarren [Clarren et al. 1987] have shown the clinical utility of a method which identifies triads of points on frontal and lateral photographs of the face and compares these groupings between presumably normal control subjects and those thought to have specific diagnoses. (Figures 3.3 a,b)



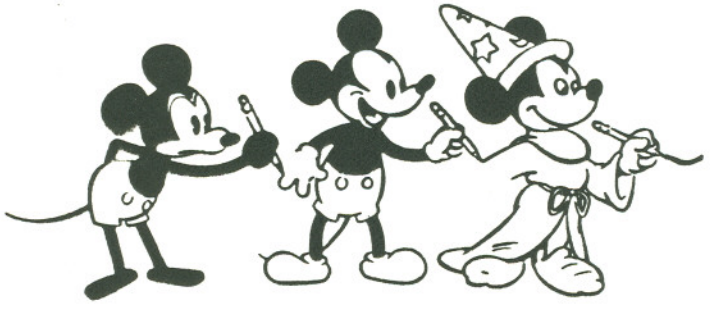
Twenty-three facial points or landmarks were selected for the morphometric study. The points were chosen because they could be reliably identified in the photographs and because they helped delineate most facial landmarks usually scrutinized by a clinician. Definitions of the 23 landmarks are given in Table III.



Shape coordinates for the triangle with vertices 14 (nasion), 19, and 22 (gnathion), and baseline determined by vertices 14 and 22. The shape coordinates are computed as the cartesian (x,y) coordinates of point 19 relative to point 22 fixed at (0,0) and point 14 fixed at (1,0).

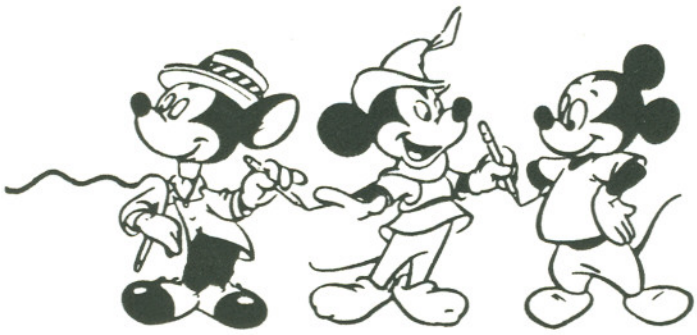
Figure 3.3 a,b. Methods of morphometric facial shape analysis [Clarren]

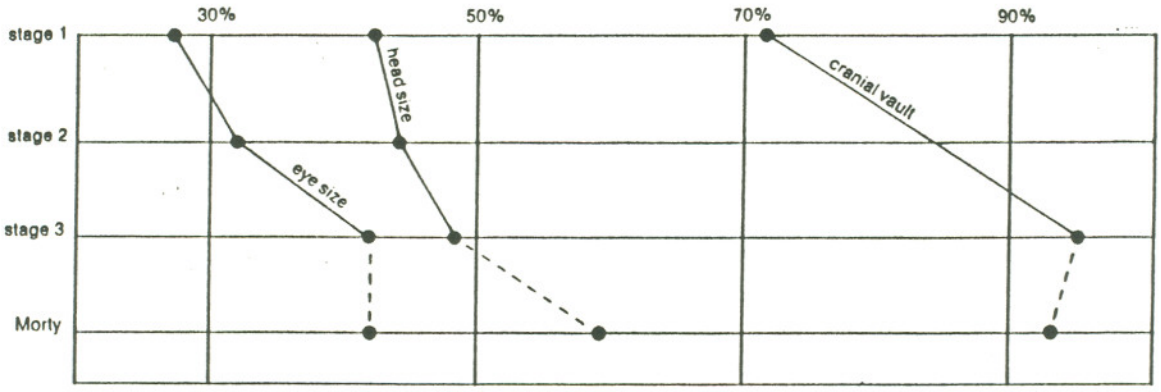
While this approach has some apparent validity for assessing craniofacial shape, it would seem to have major theoretical shortcomings as a generalized method for assessment unless a few significant assumptions are made. The human face is, for the most part, a smooth surface with a relatively small number of distinctive landmarks. In theory then, any representation of the face has potentially an infinite number of points which might be selected for measurement. The problem of which point(s) to select is probably not a trivial undertaking and may be very susceptible to observer bias. For example, extensive research has demonstrated that humans have predetermined responses to different facial configurations. When caricatures are presented to infants, if certain elements of a face are present, the infant generalizes and responds as though the entire face is present. In older children and adults, variations in facial configuration result in strong differences in emotional responses. These data suggest that what we see when we look at a face may be determined to some extent by the effect of genetically-determined characteristics of our nervous systems and postnatally acquired experiences. This is discussed rather humorously by Gould in his paper on the evolution of the facial structure of Mickey Mouse. [Gould 1980] (Figures 3.4 a,b,c)



Mickey's evolution during 50 years (left to right). As Mickey became increasingly well behaved over the years, his appearance became more youthful. Measurements of three stages in his development revealed a larger relative head size, larger eyes, and an enlarged cranium—all traits of juvenility. © Walt Disney Productions

Figure 3.4 a. The evolution of Mickey Mouse. [Gould 1980]





At an early stage in his evolution, Mickey had a smaller head, cranial vault, and eyes. He evolved toward the characteristics of his young nephew Morty (connected to Mickey by a dotted line).

Figures 3.4 b,c. The evolution of Mickey Mouse. [Gould 1980]

Therefore the points which might be selected for the application of this, or related, approaches might not always be chosen objectively. In addition to this potential problem, a serious difficulty is inherent in Bookstein's method. With the exception of a relatively small number of landmarks, only a few locations can be identified reliably on photographs of the face. Large areas, which may or may not be of importance in a given disorder, are not really identifiable because they lack detail. The cheeks, for example, while frequently described as being either full or flat in clinical descriptions are in either case smooth and do not generally have reliable variations in pigmentation upon which to fix a point. So the points selected for this method must obviously be chosen because it is possible to find them, not necessarily because they are the best points to use.

Consider also the problem of how many points are required to represent a face adequately for diagnostic purposes. In some respects, this is a chicken and egg type of issue. There is probably no system now in existence under which trials could be undertaken in order to test any hypotheses, and one needs to have some estimate of minimum resolution in order to design such a system. One possible approach to this dilemma is to use a dysmorphologist as the "system" and present images of varying degrees of coarseness. This approach assumes, of course, that the human system is not significantly better or worse than a hypothetical computerized system with regard to extracting meaning from an image and also that a two-dimensional image is an adequate representation of a three-dimensional subject. Accepting the limitations of these assumptions, a very simplistic test model of the maximum degree of image coarseness was constructed using digitized images of children with a known syndrome. The initial digitization was at a resolution of 512 by 512 pixels with 256 shades of grey. These digitized images were then resampled at varying degrees of resolution down to 16 by 16 pixels. As shown below in the images in Figures 3.5a through 3.5d, to the human observer pixel arrays with dimensions from 512 down to 64 do not appear to exhibit significant degradation in recognizability. Below 64 pixels the image quality deteriorates substantially.

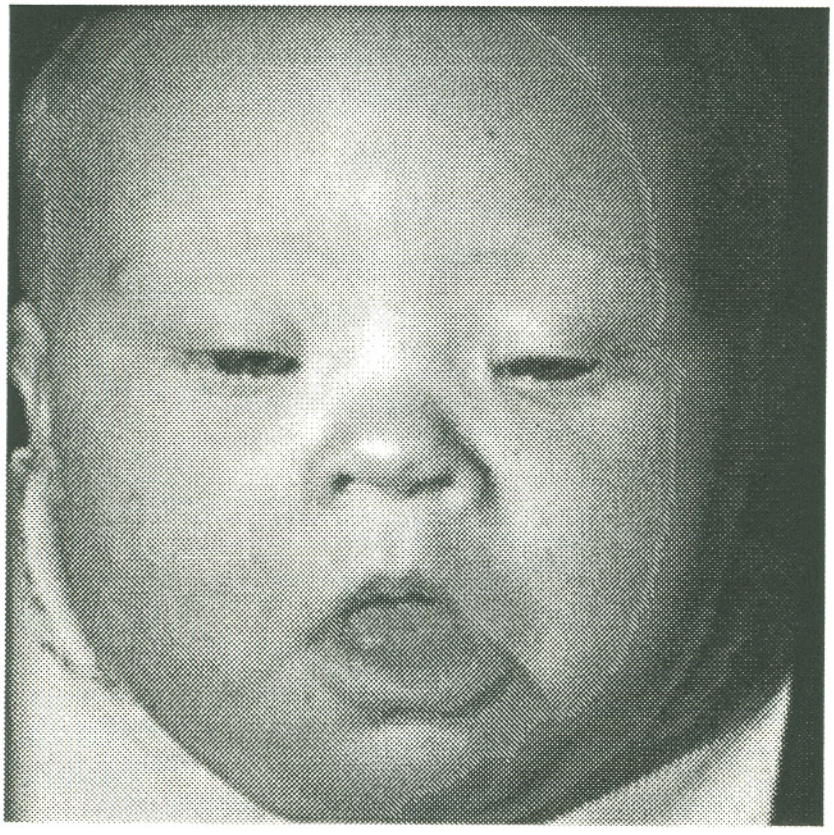


Figure 3.5 a Image of a child with the Down syndrome 256 x 256 pixels

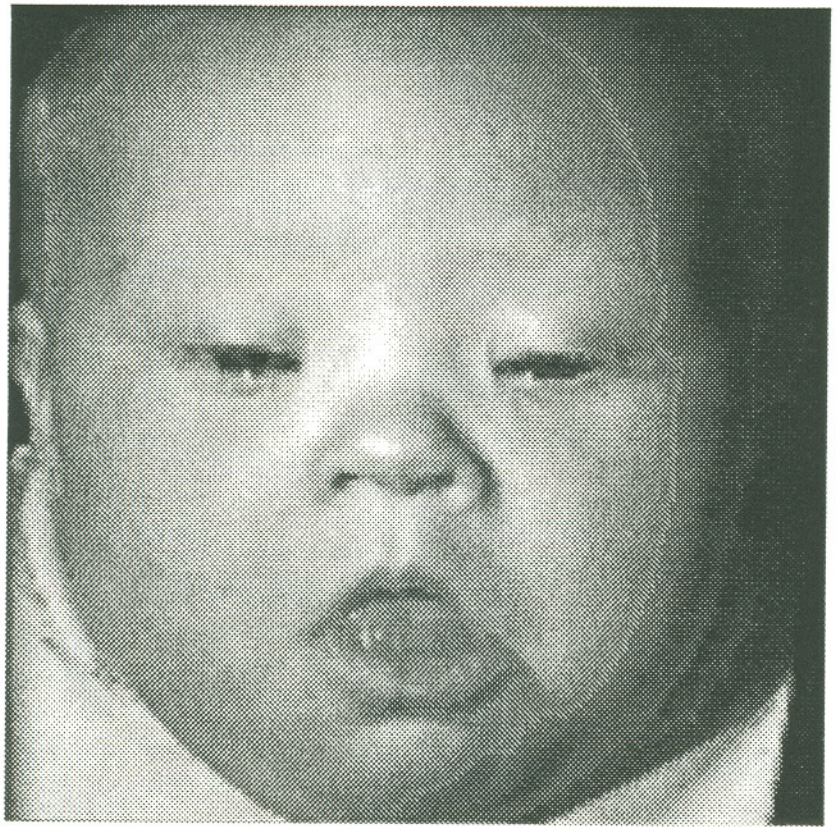


Figure 3.5 b Image of a child with the Down syndrome 128 x 128 pixels

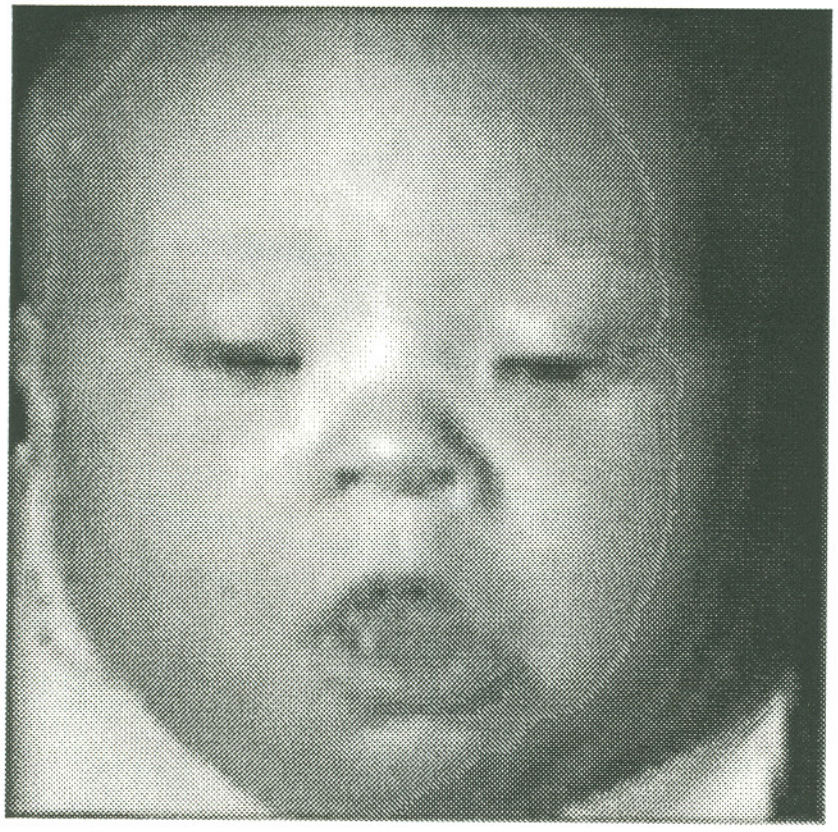


Figure 3.5 c Image of a child with the Down syndrome 64 x 64 pixels

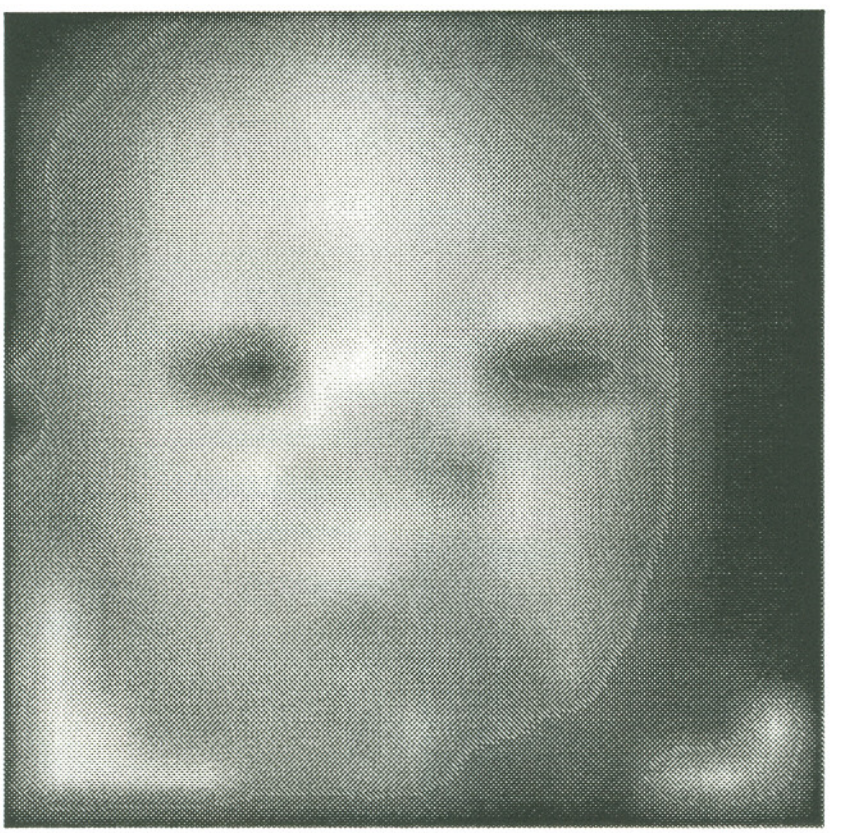


Figure 3.5 d Image of a child with the Down syndrome 32 x 32 pixels

This admittedly crude experiment would appear to place some lower limit on the number of data points needed to characterize adequately a human face for the purposes of dysmorphologic analysis, at least for 2D projections. The 64 by 64 pixel images appear to be nearly satisfactory, with some obvious loss of information for the clinician, while lower resolution images show signs of degradation and the 16 by 16 pixel images are not recognizable. In terms of image scale, an adult human face is perhaps 200 mm high and 120 mm wide so an image grid of 2 - 4 mm seems to be the maximum degree of usable sampling coarseness - at least by the standard of 2D photographic images. Informal discussions with researchers working on 3D reconstructions of the head also suggest about a 2 mm maximum sampling coarseness, with 1 mm as an ideal.

CHAPTER 4

TWO-DIMENSIONAL VERSUS 3-DIMENSIONAL IMAGING

The development of a system to aid in the objective evaluation of the human face could use either 2 dimensional images or 3 dimensional representations. It has been shown that measurements obtained systematically from 2 dimensional photographs can be used, in at least some clinical settings, to separate a clinical diagnosis based on facial features. It is not yet clear whether this approach can be generalized to all situations requiring objective assessment of craniofacial configurations. Clinical intuition suggests that if all craniofacial diagnoses could be made objectively from 2 dimensional projections of facial configuration, i.e. photographs, then clinicians would not feel the need to see patients in person and would be comfortable working from slides or prints on a regular basis. At times diagnoses may be quite obvious from even less than optimal photographic images, but not infrequently even high quality images fail to convey the clinical message. This may be a question of degree. Where the craniofacial configuration is substantially altered, the 2 dimensional projections will reflect the degree of alteration. When the alterations are more subtle, the projection from three to two dimensions may cause enough loss of information to limit usefulness. It is also possible that sufficient information remains in the 2 dimensional images but clinicians are not facile at extracting it. In this case some variety of computer-assisted methodology might be of value.

One approach to computer-assisted analysis of shape involves the use of Fourier descriptors. [Gonzalez & Woods 1992] A variety of specific methodologies have been used, but they generally involve the conversion of a parametric representation of a curve into a small set of "descriptors" which are coefficients of a Fourier (or related) transform and characterize the shape of interest in a more compact form. The descriptors therefore represent frequency and phase coefficients for the transform. While this method has some merit in terms of representation and perhaps analysis of biologic shape, it does not seem to provide any useful insight, particularly the clinically intuitive type, into understanding an entire face. [DiLiberti 1986] A relatively small change in one of the coefficients could

substantially alter the represented shape, yet an examination the numbers would give a clinician no understanding how the resulting shape would appear and what its clinical/anatomic meaning might be.

CHAPTER 5

MAGNETIC RESONANCE AND COMPUTERIZED TOMOGRAPHIC SURFACE RECONSTRUCTIONS

Two powerful imaging techniques, CT (computerized tomography) scanning and magnetic resonance imaging (MRI) have revolutionized medical diagnosis during the past 2 decades. Although used primarily for evaluation of internal organs and body structures, each has the potential to generate data for surface reconstructions. (Figure 5.1) Both modalities produce sequential 2D grey scale cross-sections of anatomic regions selected for study. The surface contours of multiple 2D sections may be reconstructed, using well-described techniques, to produce a 3D representation. [Hoehne et al. 1987] (Figure 5.2)

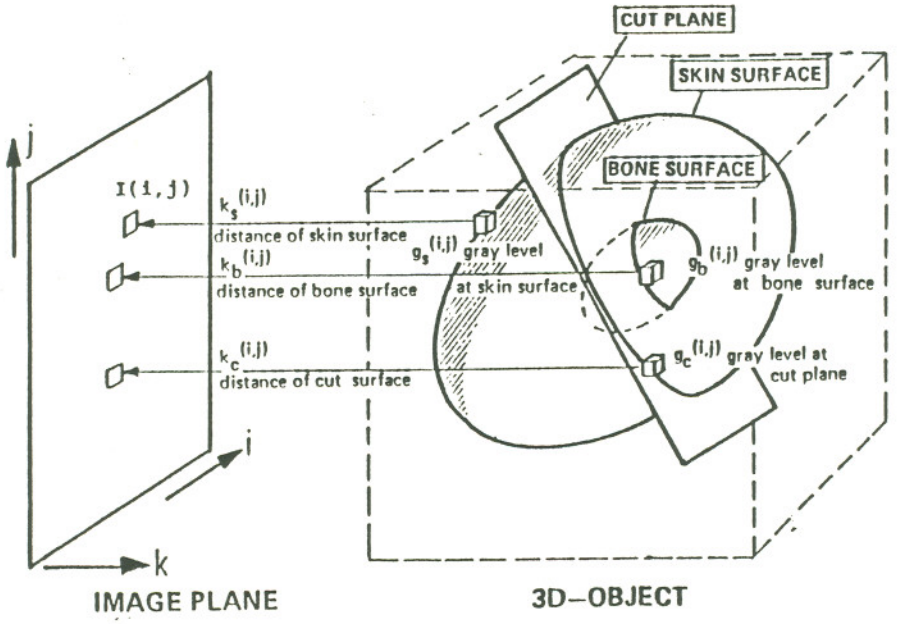


Figure 5.1 Geometry for tomographic surface reconstruction. [Hoehne et al. 1987]

Figure 5.2 Surface reconstruction of the face from tomographic images.

The accuracy of this 3D representation compared to the actual surface in question is limited by 3 main considerations. First, it is obvious that the distance between successive 2D sections or "slices" limits the theoretical resolution along the axis perpendicular to the slices. Thus, accuracy on the order of a few millimeters requires slices at similar intervals. Second, limitations in the reconstructions algorithms along with typical hardware specifications of 256 x 256 to 512 x 512 pixel maximum image resolution may influence accuracy of surface reconstructions to some extent. Finally, despite the fact that scanning time in modern equipment has been greatly reduced, the possibility of patient movement must be considered - particularly in children.

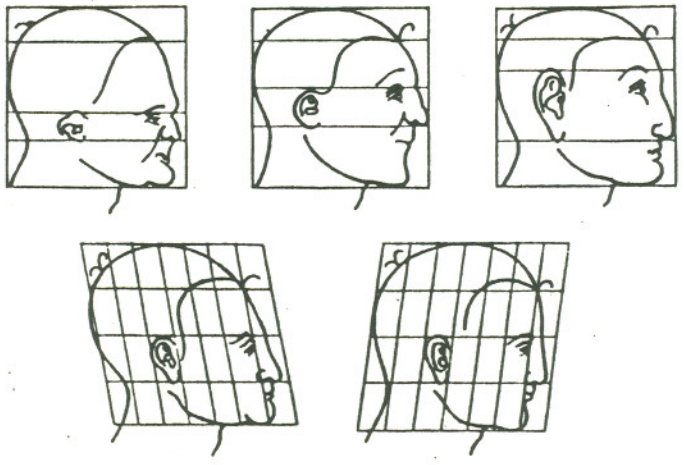
These issues regarding accuracy are probably of relatively minor significance compared to concerns about cost, safety, and inconvenience to patients and physicians. Routine MRI and CT scan charges are typically on the order of several hundred dollars, but require far fewer slices than would be necessary for a 1-2 mm resolution scan series for surface reconstruction. It would be more difficult to obtain scans routinely since patients would have to be scheduled in advance for the procedure and transported, perhaps some distance, to the MRI or CT scanning facility. Finally, the radiation dose for CT scanning would be significant considering the number of slices that would be necessary. Although there are no known risks for MRI scanning, the duration of the procedure would almost certainly require fairly deep sedation, in children at least, with the attendant potential for complications.

Scanning times for MRI scanning will probably continue to fall, perhaps into a range that would allow for use of this technique for routine surface reconstruction. In the meantime, except in circumstances where perhaps it is important to have surface reconstructions along with the relationships of the skin surface to the underlying soft tissue and bony structures, it is unlikely that MRI or CT scans will achieve routine use in this field.

CHAPTER 6

OTHER MISCELLANEOUS APPROACHES

One of the earliest approaches to the mathematical assessment of biological shape was the use of coordinate transformations by D'Arcy Thompson early in this century. [Thompson 1961] Bookstein's approach to the analysis of the human face apparently borrows substantially from Thompson's ideas. Thompson proposed, in essence, that if the shape of an organism is plotted out onto Cartesian coordinate axes, then the transformations of the axes needed to map the shape of this organism into the shape of another organism provide some insight into the biological transformations (the evolutionary steps for all intents and purposes for transformations between species) which would transform the first organism into the second. Thompson borrows this concept from some much earlier work by Durer who was interested in the geometric transformations needed to convert one facial type into another. (Figures 6.1 a,b)



(After Albrecht Dürer.)

Figure 6.1 a. Coordinate transformations of the face. [Thompson 1961]

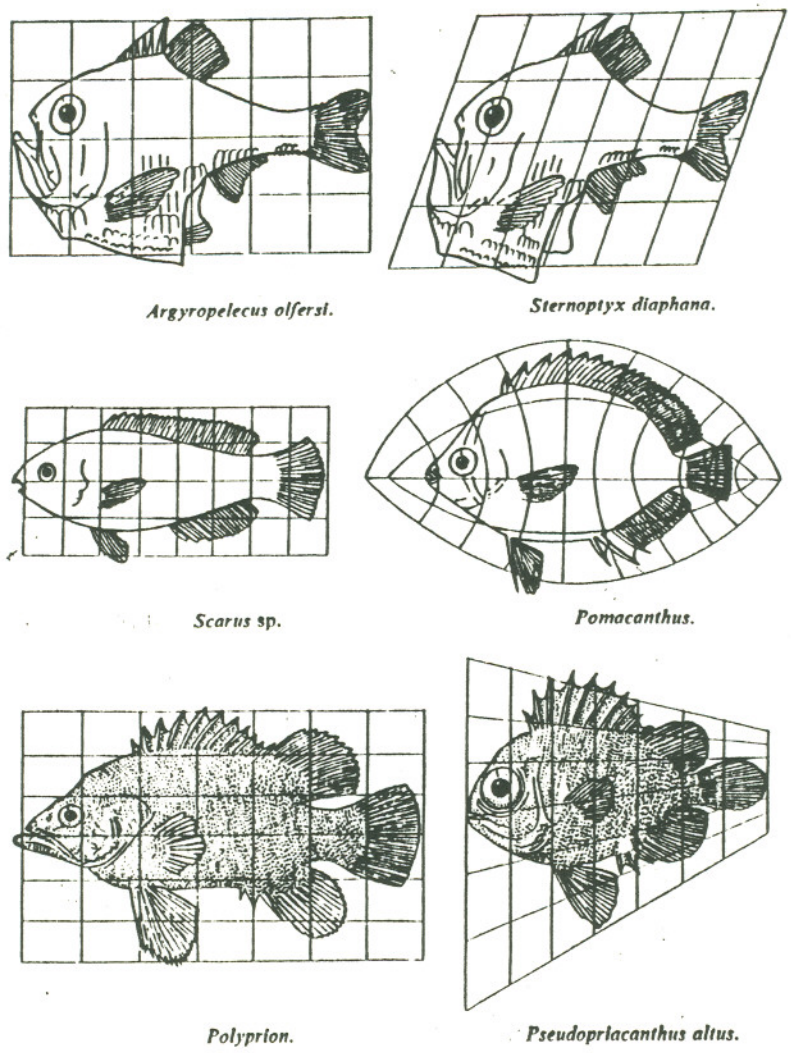


Figure 6.1 b. Coordinate transformations of fish species. [Thompson 1961]

While Thompson's notion with regard to transformations between species appears to be very naive based on our current knowledge of evolution in the late 20th century, at the time it was probably very sensible. If one steps back from the evolutionary implications and looks at this method purely from its potential utility in the assessment of the differences between shapes rather than imputing any biological meaning to the differences in the coordinate transformations, then the method becomes more interesting.

Accordingly a system was designed for the use of coordinate transformations on 512 x 512 pixel grey-scale photographs by the method of grid-warping. While it was clearly feasible to transform a 2 dimensional image, for example from a normal face to the appearance of a known condition, when presented to an audience of dysmorphologists and clinical geneticists there was less than overwhelming enthusiasm because the method did not seem to offer them any additional insight into understanding the difference between shapes. [DiLiberti 1990; DiLiberti 1991] (Figures 6.2 a,b,c) This may have been a reflection of the constraint that it was a 2 dimensional system. In 3 dimensions the comparison between surfaces may have been far more helpful than the 2 dimensional comparison of projected images.

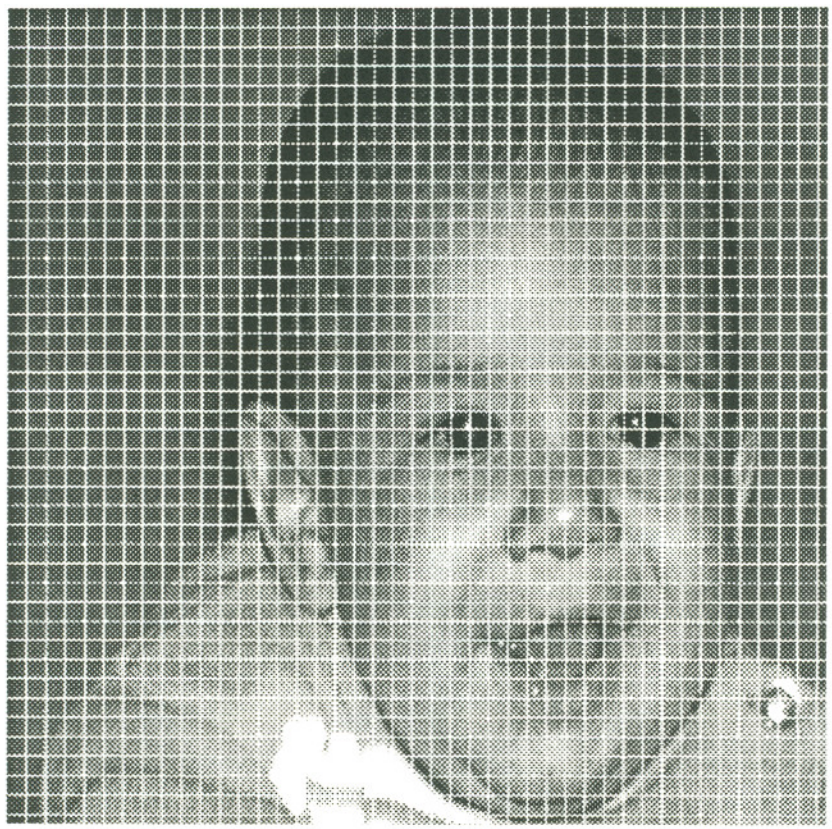


Figure 6.2 a Image of normal child with superimposed grid

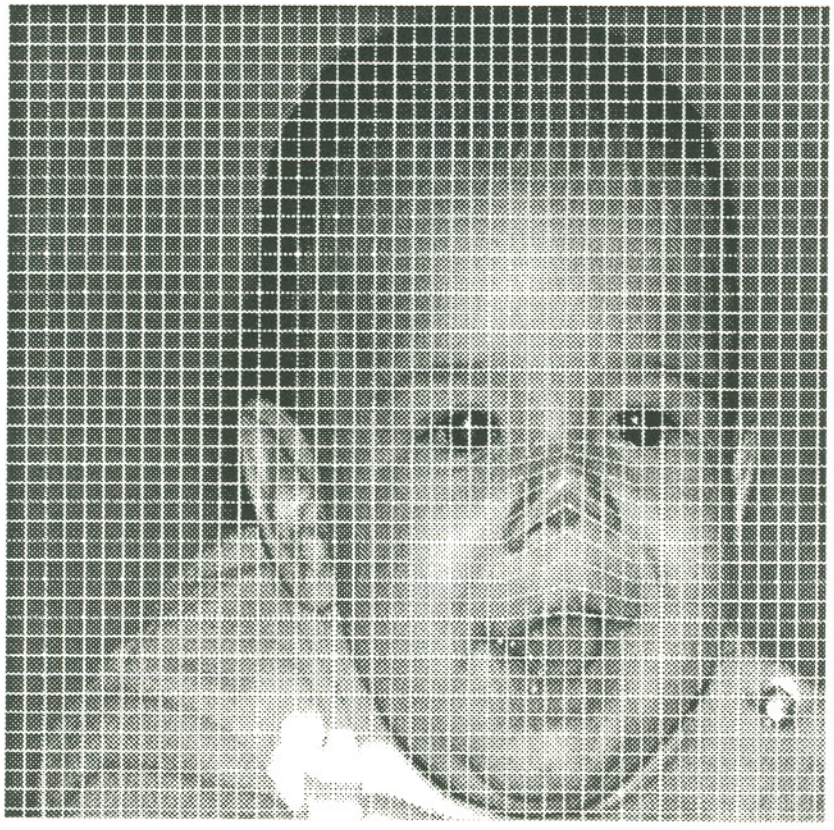


Figure 6.2 b Same face as 6.2a with single-point coordinate transformation

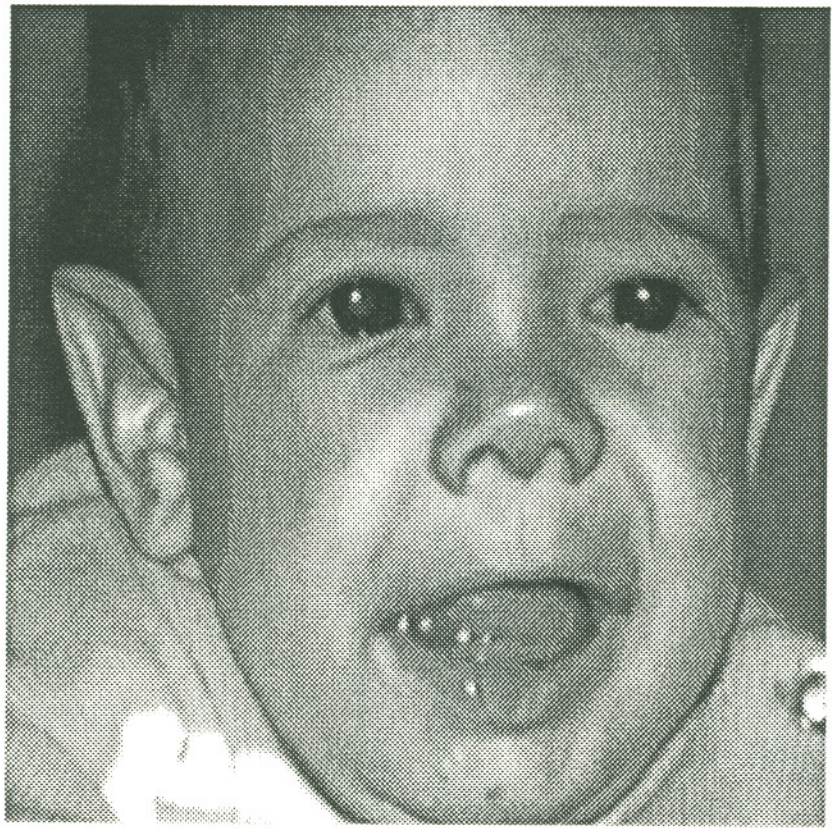


Figure 6.2 c Same face as 6.2 a with multiple transformations

Holography was considered as a possible method for obtaining 3 dimensional contours of the face but discarded for several reasons. First, the optical requirements would appear to be too stringent to move the equipment into many clinical settings. Vibration will seriously degrade holographic images, which makes any sort of hand-held device difficult to design and use. Second, although the risks of low power lasers are minimal, obtaining consent from parents for infants and children could be problematic

Moiré photography is likewise rather cumbersome in its equipment requirements, making it rather difficult to imagine a unit which could be used in a newborn intensive care nursery, for example. An additional problem is that the Moiré band patterns give relative, rather than absolute, measurements in the z axis. Although there are potential solutions to this latter problem the combination of shortcomings makes it unlikely that the Moiré method is of significant promise for this application. [Hojo et al. 1982]

Although it is possible to obtain an approximation of a third dimension using various shape from shading algorithms, these again give relative values for the z axis and would tend to have substantial difficulties with areas of the face which have extensive variations in pigmentation such as eyebrows, lips, etc.

One clinical situation where objective assessment of facial configuration would be quite useful is the management of pain in young children. Unable to talk well enough to express their discomfort, they are frequently undermedicated to some extent because their caregivers are not certain whether they are experiencing pain or are fussy for some other reason. It has been shown that there are certain characteristics of the grimace a child makes in response to pain which may be used as objective criteria to assess pain and which differ from other, perhaps superficially similar, facial expressions. [Grunau & Craig 1987] (Figure 2.1) Unfortunately these objective assessments have required frame by frame analysis of video tapes by highly trained observers, greatly limiting their usefulness except in highly structured research settings. This seemed, however, to be a good problem to begin to test the effectiveness of a neural network for the evaluation of craniofacial configuration. [Rumelhart 1986]

Rather than starting out with children experiencing pain, a difficult clinical research situation, it seemed that using adult volunteers with an elicited stereotypic

response, a "happy" or "sad" facial expression[#] would be more manageable. Accordingly, a total of 68 separate images of 5 adults who were asked to present happy or sad facial expressions were digitized (34 sad faces and 34 happy faces) and used as a training set for a three-layer, feed forward neural network using the back propagation algorithm. The training set showed rapid convergence and the network was then presented with 20 images from new volunteers, not part of the original training set. In this latter test the network achieved a 100% success rate differentiating smiling from sad facial expressions. The next step in this project will be to test this concept on children scheduled to undergo painful procedures to see to what degree the neural net may be able to identify the facial expression characteristic of pain in infants. [Donohue et al. 1991]

Although the neural net approach to identifying characteristic craniofacial configurations appears to show some promise based on the simple experiment noted above, extrapolation of the results to genetic or dysmorphic conditions may not be straightforward. For one thing, some of these conditions are so rare that obtaining a training set might not be realistic - particularly if the net would need to be trained on different sets for various age ranges and perhaps different genders. In addition, race and ethnicity might be additional parameters to consider. Finally, a neural net will, in general, give an all or none response. The clinician might be more interested in the degree of similarity, or perhaps the anatomical regions where there are similarities or differences.

A neural net might prove to be more useful when dealing with 3 dimensional representations of dysmorphic craniofacial configurations. If the 3 dimensional representation is divided up into regions of anatomical interest, a series of neural nets might be trained to deal with each region. For example, the dysmorphic conditions might be viewed as structures built up from simpler pieces. There is, in fact, some embryologic truth to this suggestion. A certain configuration of the lips might not be specific for one condition, but might appear in a variety of conditions, including normal populations. It is the association of a specific configuration of the lips with specific configurations for other anatomical regions which leads to a clinical diagnosis. This might be viewed as somewhat analogous to parsing, which also amounts to the identification of simpler parts (and the way in which they are used) from which a more complex structure can be built. Of course there are other statistical ways in which the surface configurations of fairly restricted

---

[#] (obviously this is a somewhat artificial situation since these facial expressions were presumably "forced" to some extent and may or may not have reflected the subjects underlying emotional state)

anatomical regions might also be compared. These approaches would be more amenable to identifying degrees of similarity which might be of considerable benefit.

CHAPTER 7

STEREOMETRIC IMAGING

At first glance stereometric imaging would appear to be well suited for obtaining 3-dimensional surface data for the human face. [Gonzalez & Woods 1992] It has been used successfully in a variety of other domains and it is reasonably practical to obtain a pair of stereophotographic images in typical clinical settings. Using a set of matched cameras and lenses mounted on a frame to maintain a fixed base between the cameras and also to assure rigidly-maintained orientation of the optical axes, photographic images have been input into traditional opticomechanical aerial mapping devices to obtain isocontours of facial surface elevations to assist in surgical reconstruction of craniofacial anomalies. Although these mapping devices seem to work fairly well, they were designed for aerial stereometric imaging and apparently no data are available examining the accuracy and reproducibility of the isocontours for images of the human face. They may also be too time-intensive to be very practical clinically . [Savara 1985] (Figure 7.1 a,b)

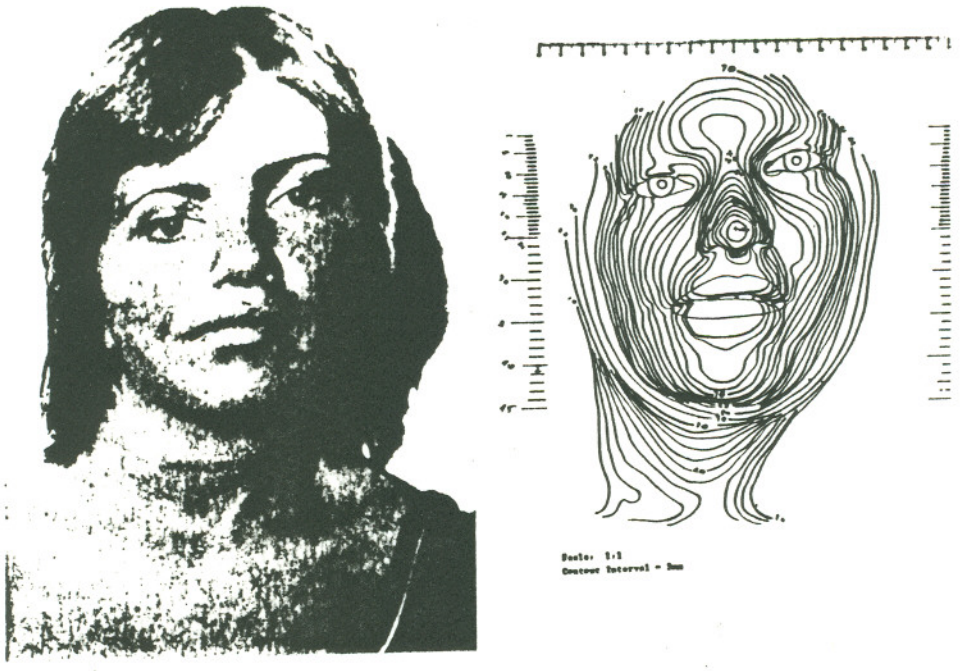


Figure 7.1 a. Surface contours of face pre-operatively

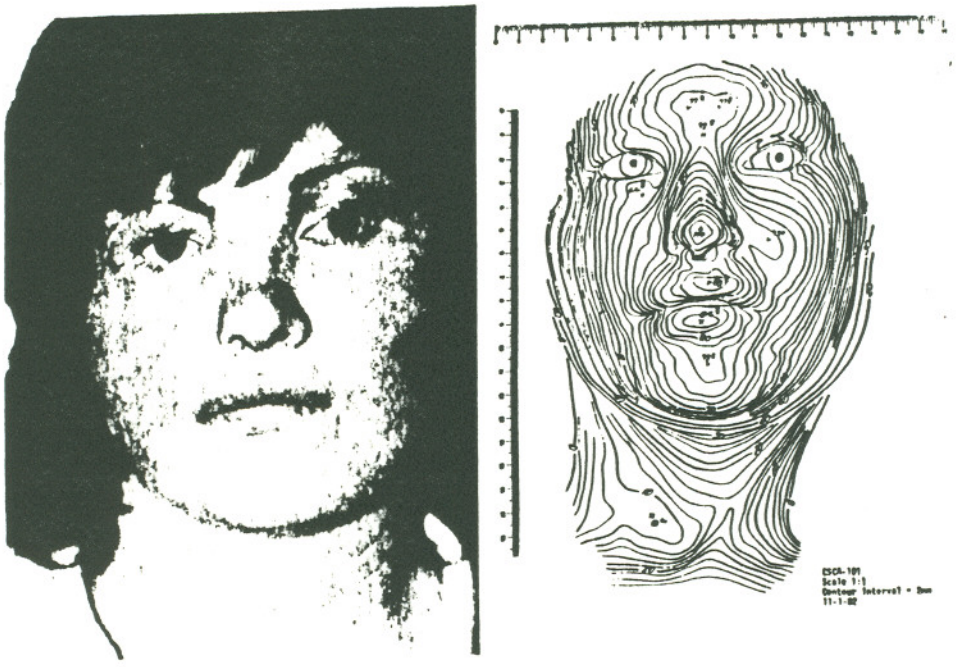


Figure 7.1 b. Contours from same patient post-operatively

Several potential problems include the expense and apparently slow turnaround time for the plotting device and the need to use two cameras. In addition, potential resolution issues, discussed below, might be of concern. For non-medical domains digitized stereometric images have been used to obtain xyz scene coordinates using computerized matching algorithms. [Yakimovsky & Cunningham 1978] This approach would have the advantages of relatively low cost and rapid turnaround time.

Prior to attempts to implement an actual working system, some preliminary calculations of the estimated resolving power in the z, or depth, axis of a practical configuration were made. In addition, similar calculations as well as a few simple experiments were performed to look at resolution in the x and y coordinate axes. Figure 7.2 shows a schematic diagram of the geometry of a typical stereometric imaging system. [Winston 1984]

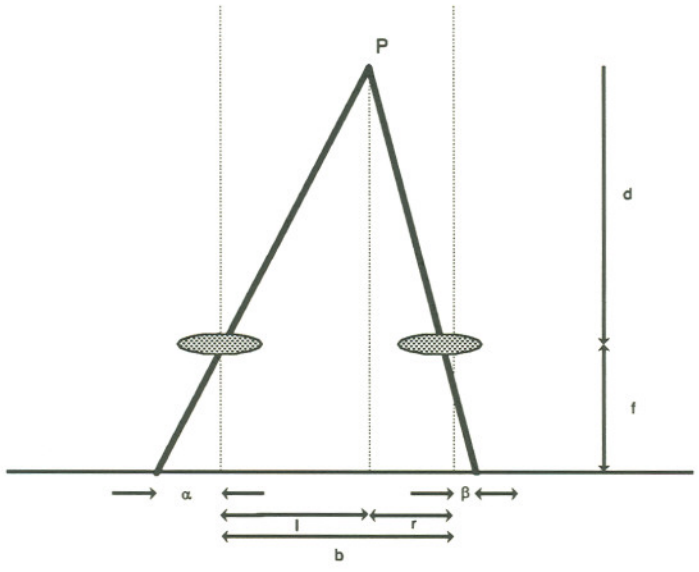


Figure 7.2 The geometry of stereometric imaging

The distance, **d**, of a point, **P**, from the plane of the optical centers of the lenses may be calculated from the image plane disparities, **alpha** and **beta**, if the geometric parameters **b**, the base or distance between the optical axes of the two systems, and **f**, the

distance between the lens node and the film plane are known. The following formula applies:

d = f*b / (alpha + beta)

Assuming a distance of approximately 100mm for the lens to film plane parameter for a typical single lens reflex camera telephoto lens, a distance of 1 meter between the subject and camera, and a 200mm base, the sum of stereo disparities may be calculated:

(alpha + beta) = f*b / d or 100mm * 200mm / 1000mm = 20mm

Note that the horizontal dimension of the film along with the image size and magnification place a limit on the maximum disparity that can be imaged. 35 mm film has a horizontal dimension of 36 mm so that the sum of the stereo disparity and image size cannot exceed this figure. With a stereo disparity of 20 mm, this leaves 16 mm for the image. At a magnification ratio of about 1:10, the maximum subject size (in the horizontal dimension) would thus be about 160 mm - the approximate width of an adult face - excluding ears and hair.

To achieve the requisite resolution in the z (or depth) axis we must be able to detect the stereo disparity of an adjacent point 1mm closer to the camera. Using the formula above, the disparity for this point becomes 19.98mm or a 0.02 mm difference in stereo disparities between the adjacent points. The requirement that the optical system have the ability to resolve at the level of 0.02mm at the film plane is within reach of high-quality lenses with fine grain, high resolution film in the center of the optical axis, but approaches the limits off-axis. [Kodak 1984] Obviously the left and right images cannot both be centered optically since the very nature of the configuration implies that at least one of the images will be off-axis. If the required resolution in the z-axis is relaxed to 2mm, under most circumstances this will be achievable.

For the x and y axes the achievable resolution will be determined more by the presence of distortion in the optical systems than film-plane resolution of the lens-film configuration since if a face, perhaps 200mm high is projected on about 20mm of the 24mm high film frame, an image resolution of 0.01 to 0.02mm on film yields a subject resolution of 0.1 to 0.2 mm by simple geometry. Measurement errors due to varying

magnification ratios are of far greater concern. Using a pair of ruled test targets and a variety of different focal length lenses and camera-to-object distances, errors in photogrammetric measurements were, as predicted, found to be related predominantly to the camera-to-object distance. Since the tip of the nose is perhaps 100 to 150 mm closer to the camera than the ear is it will appear larger on the film plane. The apparent difference in magnifications is determined by the equation:

M = difference in distance from camera / distance from camera

For an object-to-camera distance of 1 meter, the maximum error for measurements in the x-y plane will be about 10%, giving an achievable overall resolution of at least 1.0 +/- 0.1 mm, which should be within an acceptable range. [DiLiberti & Olson 1991]

Accordingly, an attempt was made to implement a system to acquire xyz coordinates of the human face using a hybrid stereophotographic computerized approach. Two Nikon N2000 single lens reflex cameras with 85mm focal length f2.0 lenses were mounted on a frame which maintained parallel alignment of the camera focal planes and a fixed 200mm distance between the optical axes of the lenses. Since these cameras have electronic shutter releases in addition to the standard mechanical release, it was feasible to wire a single, hand-held shutter release switch to trigger both cameras simultaneously. A single electronic flash was attached to one of the cameras in order to have adequate lighting under all clinical situations. Fine grain black and white panchromatic film (Kodak TMax 100) was exposed and developed according to the manufacturer's instructions. It was immediately apparent upon inspection of the developed film that simultaneous electrical triggering of the cameras did not lead to perfectly synchronous shutter movement. In one of the cameras the shutter was not synchronized with the flash attached to the other camera, even though it would synchronize when used alone with the flash. Although this could have been remedied with electronic circuitry, the use of available light for the initial tests circumvented this difficulty, as did the subsequent use of flash bulbs which have a longer duration of light production than an electronic flash and hence increase the margin for error in synchronization.

Images were digitized at a resolution of 1024 by 1024 pixels with a grey scale of 256 (8 bits per pixel) using a Panasonic video camera with a Micro Nikkor 55mm macro lens, a Chorus Data Products video capture board, and a photographic enlarger for

holding the film and light source. These images were then registered in the vertical axis and standard algorithms to calculate z coordinates by disparity matching were applied to each horizontal scan line. By examining a segment of pixels in a "window" along the scan line in one image of the stereo pair, the algorithm attempts to find the corresponding region on the same scan line of the other member of the pair. [Grimson 1981] Unfortunately these attempts proved to be futile. With any reasonable size window, attempts to match a corresponding window in the opposite image got lost quite readily.

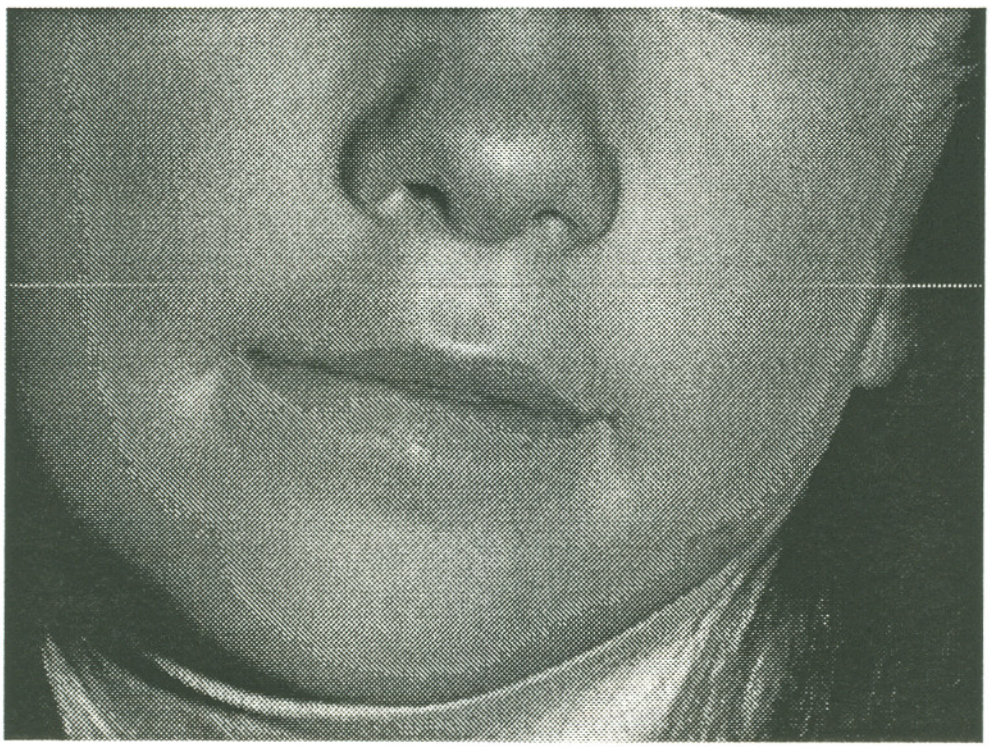


Figure 7.3 Digitized image of face with scan line highlighted

The reason for this was immediately apparent when the image files were examined more carefully. Figure 7.3 is a grey-scale image of a section of a photograph of the face of a 7 year old girl with a single scan line highlighted.

Figure 7.4 a. graphs the pixel intensities for that scan line.

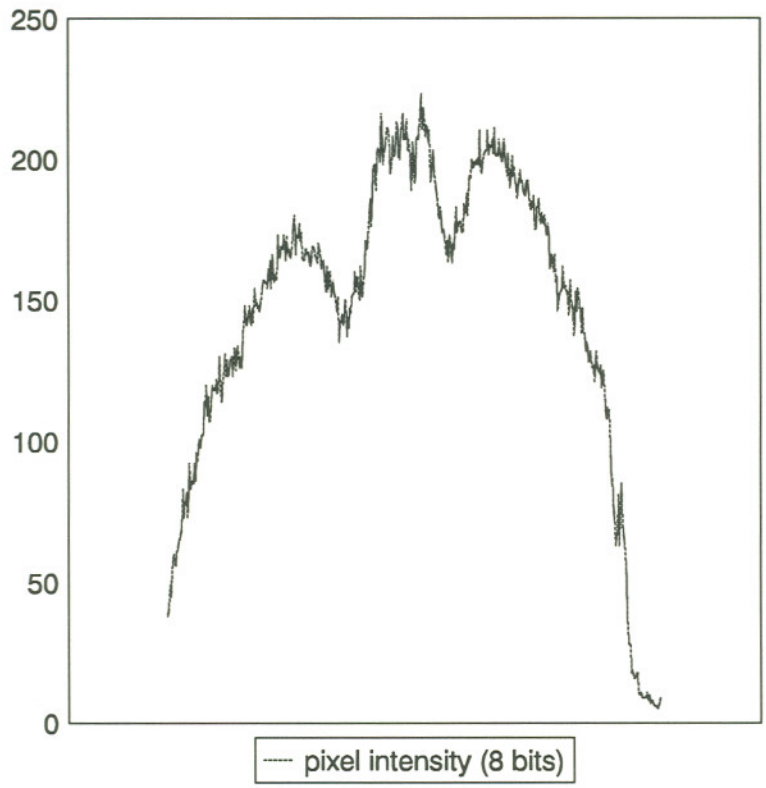


Figure 7.4 a Pixel intensities for scan line in Figure 7.3

There is obviously a lot of high-frequency noise - presumably resulting from graininess of the film and electrical noise in the video imaging system. This leads to the very jagged edges clearly visible in the figure. When this noise is filtered by convolving the original image with either a 3 x 3 or 5 x 5 averaging filter smoother images with few significant details resulted. (Figures 7.4 b,c)

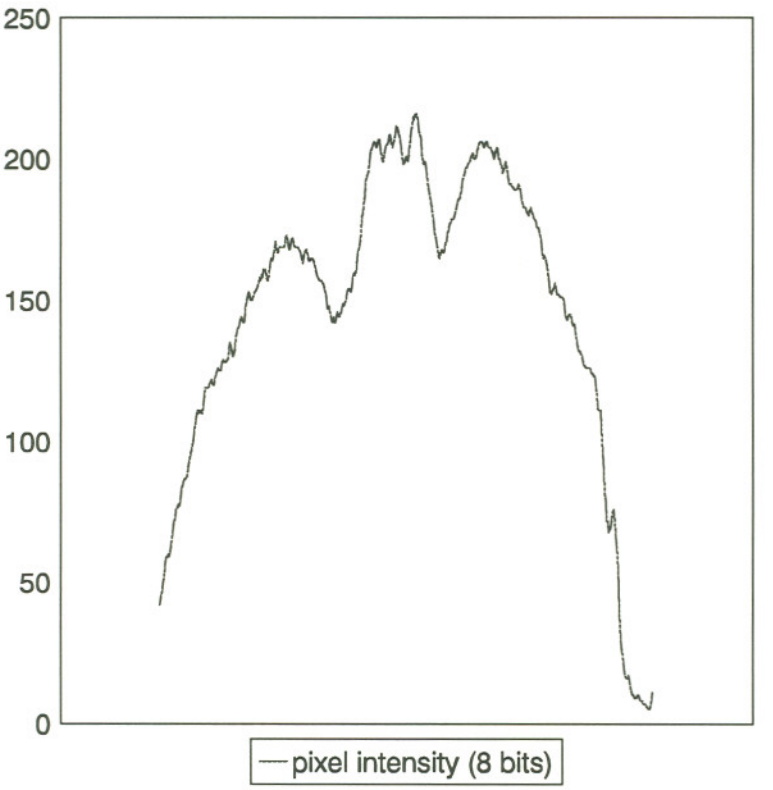


Figure 7.4 b Scan line pixel intensities with 3 x 3 smoothing

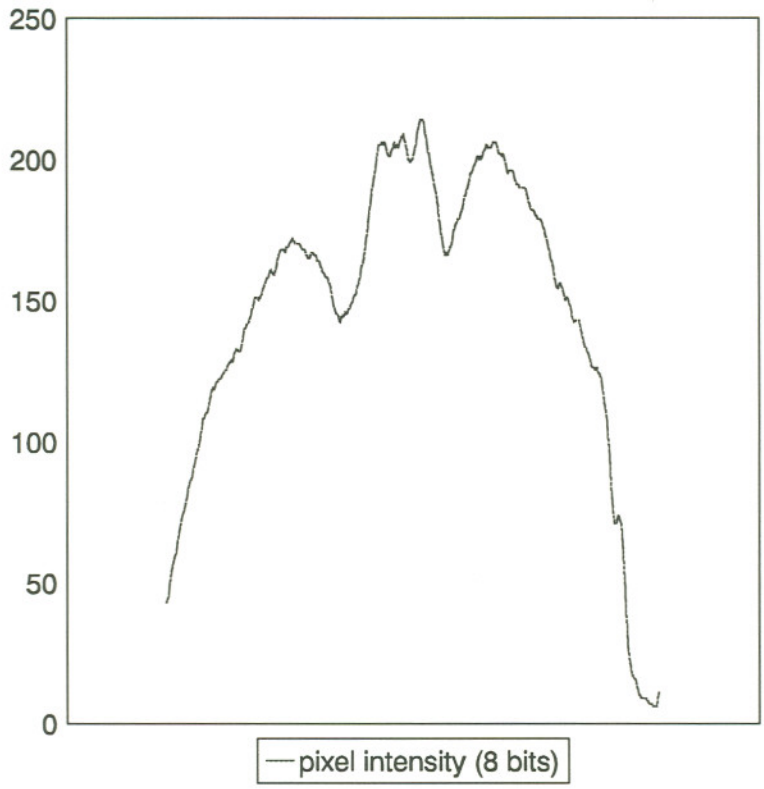


Figure 7.4 c Scan line pixel intensities with 5 x 5 smoothing

These data fit in with subjective inspection. For the most part, a human face - particularly that of a child - has very few details for point to point matching. As can readily be seen, there are really no significant fine details which could possibly be matched to the corresponding stereo image. One possible solution is to project a pattern onto the face which allows identification of regions with 1 mm resolution in corresponding images. While perhaps theoretically possible, the practicality of finding correspondence is challenging. If for example, vertical, regularly-spaced black and white bands are projected, the resulting images have high contrast patterns for matching. Deciding which bands

match in the corresponding images becomes quite difficult. Since each image is taken from a different position, each camera has a slightly different angle of view of the face. One image will therefore have greater coverage of the right side of the face and the other greater coverage of the left. The bands on the leftmost side of the face in one image will not be visible in the other image - the opposite is true for the rightmost bands. Thus it is not possible to start the matching algorithm at the first visible band on the left sides of the images since they do not correspond to the same projected band. Using one uniquely identifiable (wider?) band - perhaps in the middle of the image - could potentially work if it is possible to count reliably every band on either side. When this approach was attempted it became apparent that some areas of the face such as nostrils, eyebrows, lips, and any other dark regions cause enough reduction in contrast that it is difficult to follow and extract all of the band patterns reliably. This method was therefore abandoned as a primary method of obtaining the required data.

CHAPTER 8

STRUCTURED LIGHT

The contours of a surface may be assessed by actively illuminating the surface with one or more beams of light and recording the resulting pattern for immediate or subsequent analysis. [Cohen & Feigenbaum 1982; Li et al. 1990] These may be grouped under the general heading of structured light since they all depend upon knowledge of the spatial organization, or structure, of the beam(s) in order to reconstruct the surface. This is perhaps best illustrated by examination of Figure 8.1.

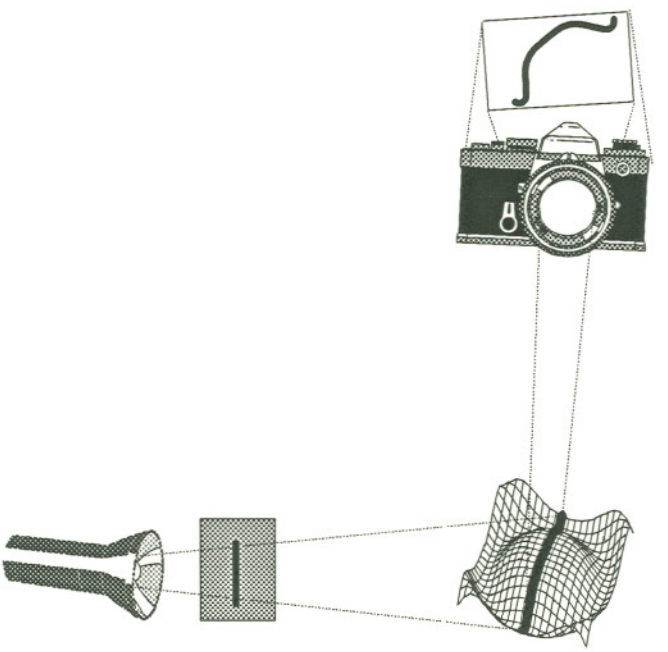


Figure 8.1 Schematic of single line structured light apparatus

An optical system projects a single narrow band of light focused on the object of interest. A camera, aimed at right angles to the projected beam, but with the film plane parallel to the plane of the band, records the light reflected from the surface of the object under illumination from the band of light. The recorded image, consisting only of light reflected from the actively illuminated area, is an isocontour for a locus of points on the object surface a fixed distance from the camera. The x and y coordinates may be taken directly from the image at the focal plane of the camera and scaled based on a readily calculated magnification ratio dependent solely on the focal length of the lens and distance from the object. Therefore, this simple apparatus gives x, y, and z coordinates for the actively-illuminated portion of the surface. It should be noted that any portion of the object surface hidden from the beam of light will not contribute to the eventual surface reconstruction. For the human face this does not present much of a problem except in a few small areas such as the region behind the nostrils for some faces. (Figure 8.2).

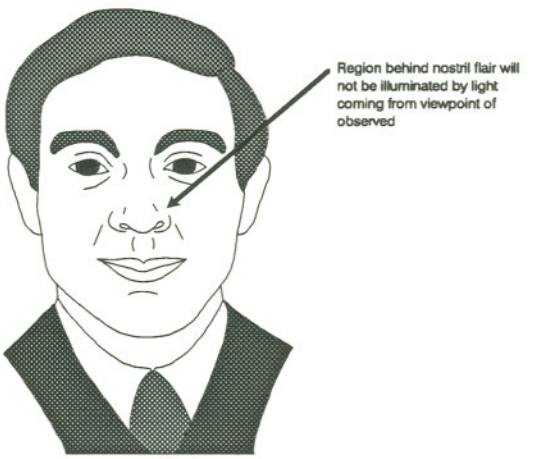


Figure 8.2 Face schematic showing hidden region

This difficulty can be overcome in almost all situations by reducing the angle between the projected beam and the axis of the camera's optical system to less than 90 degrees. This slightly increases the complexity of the calculations for reconstructing the surface coordinates and, depending upon how much the angle is reduced, decreases the resolution of the system. Obviously, if the camera and projector were aimed on the same axis, the z coordinate would not be accessible at all. Reducing the angle between the optical axes to approximately 45 degrees greatly reduces any likelihood of difficulties with

hidden surfaces, preserves adequate resolution in the z axis, and also makes it feasible to construct a hand-held apparatus with the camera and projection device rigidly mounted together. The latter property has the advantage of fixing the geometric parameters for reproducible coordinate calculations.

To obtain the entire set of surface coordinates several different strategies may be used, each representing a variation on the need to scan the entire surface. First, the object of interest can be moved through the narrow band of light, either linearly or by rotation, at a known rate. If multiple images, taken at known time intervals are obtained, the entire set of surface coordinates can be reconstructed. Alternatively, the active illumination source can be moved at a known rate and timed images again obtained, producing essentially the same result as in the first method. Obviously each of these approaches is flawed for most clinical settings due to the impracticality of maintaining a human face so precisely positioned in such controlled circumstances. The likelihood of an unacceptable amount of artifact from subject movement with either approach is extremely high. For example, even with a video imaging system operating at 30 frames per second, to obtain 1 mm resolution across a face perhaps 100 mm wide would take 100 frames or at least 3.3 seconds - an eternity for an active young child. Even for a child moving at a relatively modest velocity of 1 foot per second (0.68 miles per hour), 10.2 mm of movement would take place between two frames. A third tactic, projection of multiple parallel narrow bands of light is much more practical. A single, fixed projection device and single camera could obtain an image with known geometric parameters in a small fraction of a second (less than 0.001 if an electronic flash is used).

The achievable resolution for a system projecting multiple bands of light will be determined by the characteristics of the optical system and the width of the bands. As noted earlier, using fine grain film under optimal circumstances a high-quality camera can resolve down to about 0.01 mm in the center of the film plane, with poorer resolutions further from the optical axis. With a magnification ratio of 1:10 this results in an effective subject resolution of 0.1 mm, a factor of 10 greater than needed. Even assuming off-axis degradation of the image by a factor of 2-3 there is still adequate margin for error for measurements in the x and y axes. The z axis resolution depends almost entirely upon the width of the projected bands, the angle between the axis of the projecting lens system and the camera lens system, and the ability of the optical system to resolve them. Assuming the bands are 1 mm wide, using the same 1:10 magnification ratio gives a film plane width of

0.1 mm, again well within the 0.01 mm central resolution of an optimal camera and film combination.

In order to test this method, a 35 mm slide with parallel-ruled black and transparent bands was made by printing the band pattern on bond paper using a laser printer, photographing the pattern onto high contrast panchromatic film (Kodak Technical Pan Film), development (in Kodak Dektol developer), and finally copying the image onto lithographic film to obtain maximum contrast. The slide was then projected onto the subject using a standard 35 mm projector and the band pattern photographed using Kodak TMax 400 film. (Figure 8.3)

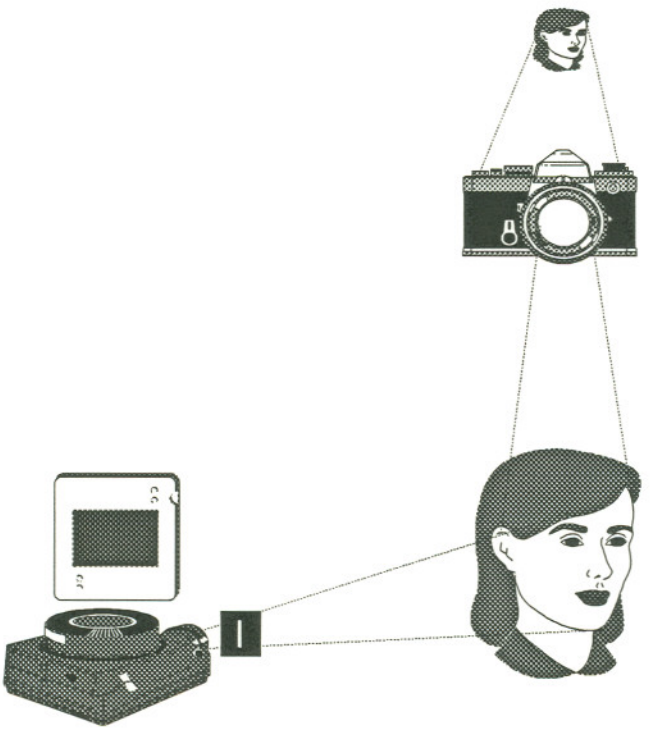


Figure 8.3 Schematic of structured light apparatus

During the initial attempts to get photographs of the band pattern projected on an inanimate object it became apparent that the bands could only be focused sharply on a fairly limited depth range - regions closer to, or further from, the projector demonstrated

blurring of the band margins. Although this difficulty could perhaps have been dealt with after digitization, since tracking the center of the band would provide the requisite depth information, the depth of sharpness could also be improved using optical methods. A standard photographic approach, readily observed in most cameras, is to use a diaphragm to increase the depth of field. (Despite the fact that taking a photograph appears to be the opposite of projecting a slide, the optical principles are the same.) Since the lenses used in almost all slide projectors lack a diaphragm, a reasonable alternative consisted of placing a small, round aperture cut out of paper centered immediately in front of the lens. While this greatly improved the depth of sharpness for the bands, as might be expected it greatly reduced the intensity of the projected light. Unlike most photographic situations where the image must be uniformly sharp, in this case the required sharpness is limited to the linear band edges. A slit diaphragm (Figure 8.3), with the axis of the slit parallel to the bands, therefore proved to be a satisfactory compromise - allowing much more light through, while at the same time maintaining edge sharpness as well as, if not better than, the round diaphragm.

The images obtained with this system revealed sharp bands across the entire test object or face, although the light intensity available from the projector on the subject was low enough to require the use of relatively high sensitivity film (Kodak TMax 400) with a normal contrast range when developed. Thus the bands appeared as shades of grey rather than black and white. If a higher light intensity had been available, a lower sensitivity, higher sharpness, high contrast film such as Kodak Technical Pan might have produced better images from which it would have been easier to extract the band patterns. (Figure 8.4)

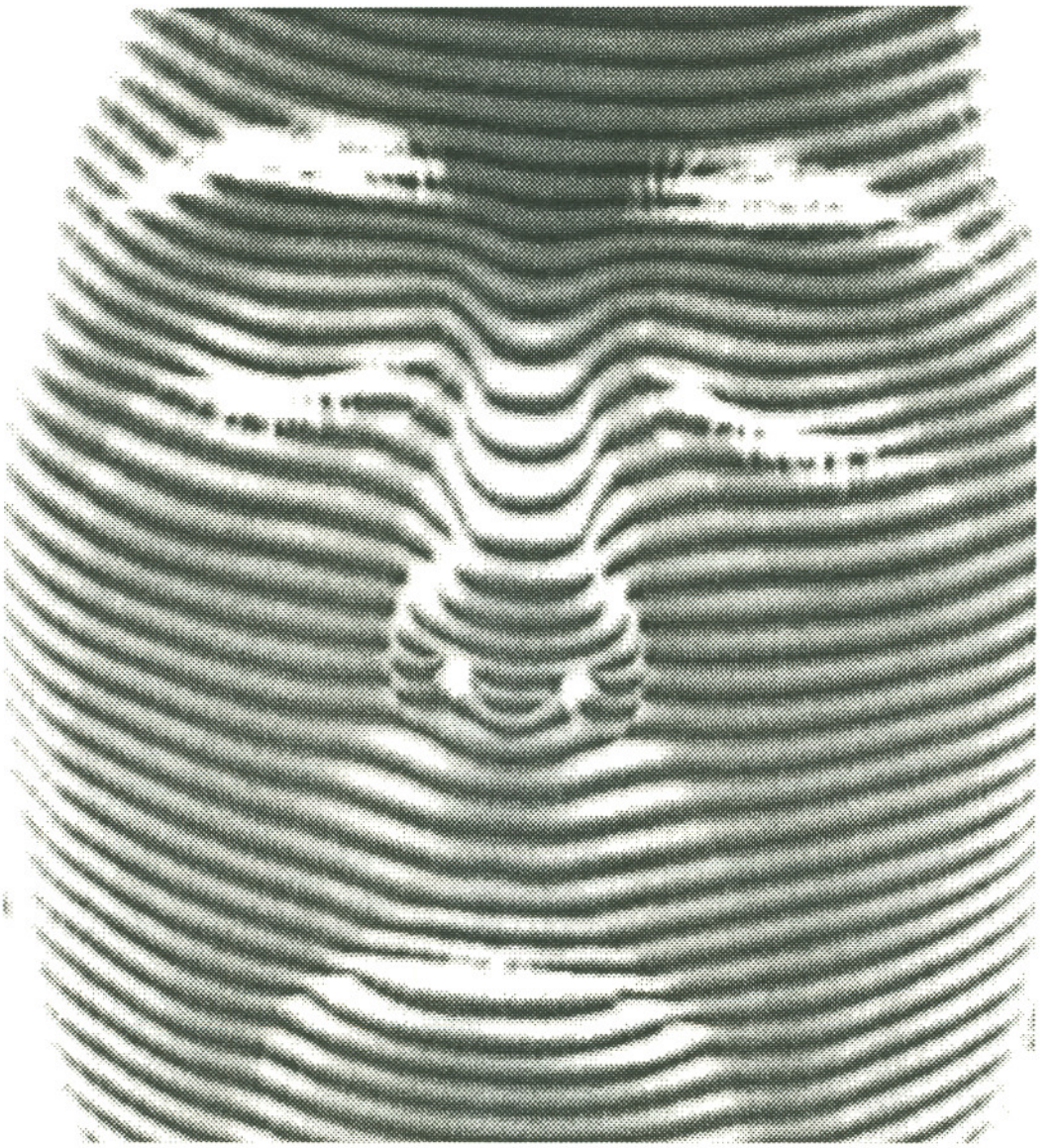


Figure 8.4 Image of face showing projected black and white bands.

The initial images were digitized at a resolution of 512 x 512 pixels with a grey scale of 8 bits per pixel (256 shades). Early attempts at extracting the band patterns from these images using a variety of standard edge detection techniques proved to be very difficult, even when contrast enhancement processing was added. The images were convolved with a number of different edge-enhancing operators including the Sobel, which seemed to give the best results. Since edge enhancement and/or detection schemes tend to increase the amount of noise in the resulting image, the relatively low contrast images had, in addition to noise from the photographic and digitization processes, "artifact" from facial details such as freckles, hair, the more heavily pigmented lip region, eyebrows. etc.

Since the images of interest are fairly unique, having a nearly regular repeating pattern in a single orientation the possibility of filtering selectively in the frequency domain arose. If a single scan line running perpendicular to the axis of the band patterns is examined by plotting the pixel intensity against position it will appear as a nearly regular, almost sinusoidal pattern. It is not a square wave because limitations in the optical system (both in the projection and camera lenses, but probably more in the former because projection lenses are not generally made to standards as high as camera lenses) degrade the highest frequency components. If instead of viewing this graph as pixel intensity versus position, it is viewed as intensity versus time, then the majority of the information content would obviously be in the (almost) sine wave with a characteristic frequency determined by the spatial positions of the projected bands. The smaller peaks and valleys in the rest of the tracing are of no interest and must be filtered out. Note that this is just the opposite of many filtering situations where the repetitive wave is the noise which must be removed. [Gonzalez & Woods 1992] Conversion of the intensity versus time to an intensity versus frequency domain is accomplished by using a Fourier transform, or closely related transforms.

The data input for a Fourier transform may be either real or complex, but the result is always complex. [Press et al. 1988] Although dealing with complex numbers is somewhat more complicated computationally than using real numbers alone, filtering in the complex frequency domain is readily accomplished. The Hartley transform is very closely related to the Fourier but has the advantage of being symmetric and operating strictly on real data, which doubles the speed of computation under some circumstances. [Bracewell 1986] By taking the scan line, transforming to the frequency domain using a Hartley transform, and examining the results, it may be seen that most of the power is in a

small range of frequencies, centered around the pattern of the projected band pattern. If the frequencies of peak power are preserved and the remainder eliminated by reducing them to zero, when the data are retransformed back into the time domain, the bands were greatly accentuated, with little or no noise enhancement.

While greatly improving the extraction of the band patterns from the images, the Hartley transform filtering was unfortunately still unable to overcome one basic difficulty. The darkest areas of the face, such as eyebrows and eyelashes, even when actively illuminated, reflect very little light to the camera and appear to be approximately as dark as the dark bands in the projected pattern. (Figure 8.5)

Figure 8.5 Band pattern extracted from Figure 8.4.

Consequently, if the brightness of these darker anatomic areas is essentially the same as for the dark bands in the pattern, neither photographic contrast manipulation nor image processing will have any significant success separating these regions. Some improvements could perhaps be obtained by eliminating all ambient light while the picture is taken. This could enhance the contrast between dark areas and bands somewhat, but is impractical in many clinical settings and would probably not solve the problem satisfactorily.

Although this appeared to be a dead-end for actively-illuminated structured light, further consideration of the problem led to a more promising approach. While the use of a black and white band pattern would apparently never be able to overcome the difficulties outlined above, increasing the complexity slightly by shifting to multispectral illumination and recording of the band pattern on color film might accomplish the necessary separation of regions. Thus a region of dark eyebrow actively illuminated with red light should be recorded on color film as dark red rather than dark grey or nearly black as on the black and white film. An adjoining region of eyebrow illuminated with green light should likewise be separable as part of the length of the green band rather than as part of the eyebrow, assuming of course that the film has enough exposure range to record dark red and dark green, etc. rather than fading into undifferentiable black.

With the entire color spectrum available for encoding the structured light the question of which colors to use and how arises. The simplest approach is to select two colors, easily separated by the imaging system, and project an array of narrow bands similar to the black and white bands described earlier except with alternating colors. This strategy should avoid all or most of the difficulty of confusing what are actually light bands on the more deeply pigmented areas of the face with dark bands. There might still be a problem tracking individual bands on the edge of the face or where there is occlusion or shadowing around the area of nose. By increasing the number of band colors it might be feasible to remove much or all of this ambiguity. If, for example, to achieve the necessary imaging resolution requires 100 to 200 bands each 1 to 2 mm in width the use of the same number of different colors as bands would uniquely identify each band. The feasibility of this approach is obviously highly dependent on the capability of the projection and imaging systems to create and resolve colors which are very close together in wavelength. An intermediate approach would be to limit the number of individual colors,

allowing for easier discrimination between colors, but still giving sufficient spatial separation between bands of the same color to eliminate all or almost all ambiguity.

In order to explore the feasibility of using multispectral structured light to obtain 3 dimensional surface information from a human face, a simple test was performed. An RGB file with parallel bands of 10 colors equally spaced across the visible spectrum was displayed on a monitor with 1024 x 768 pixel resolution and 24 bits per pixel. This image was then photographed using 35 mm color transparency film. The resulting slide was projected onto the face of a subject and another photograph taken onto color transparency film. This image clearly demonstrates well-defined bands, readily differentiated from each other, with little evidence of difficulty separating the color banding from variations in skin pigmentation. (Figure 8.6)

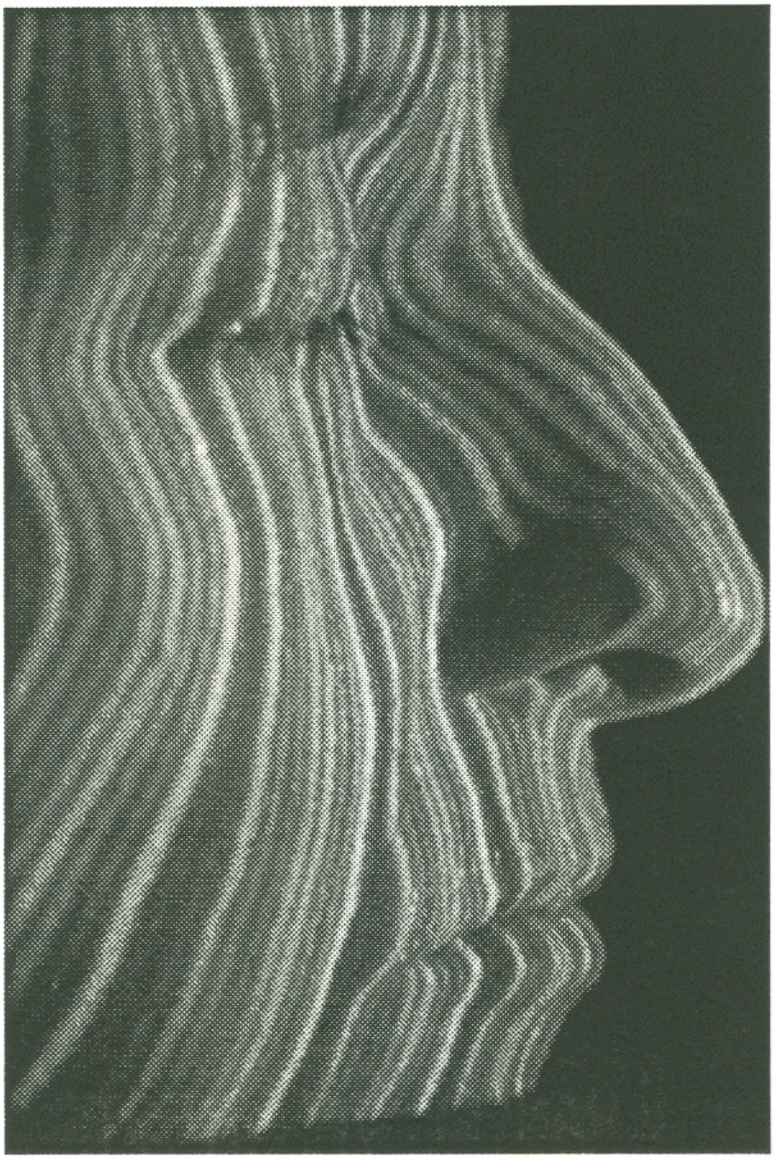


Figure 8.6 Image of face showing projection of color bands.

The only immediately obvious defect with this demonstration was the distortion of the bands in the projection slide due to the curvature of the monitor surface.

The availability of a Polaroid CI 5000 film recorder eliminated this problem. Band patterns could be output at a horizontal resolution of 4096 pixels and the flat screen of the cathode ray tube in the recorder made spatial distortion negligible. The CI 5000 has customized exposure configurations which match the spectral characteristics of a variety of films, suggesting that the colors in the final slide would be reasonably close to those in the image file. When a series of tests was run, however, by creating files of a single color, producing a slide on the film recorder and then digitizing the slide in a Nikon LS3510AF scanner, with analysis of the resulting data file for color distribution, the results were worse than expected. About 40 different colors, equally spaced across the visible spectrum, were tested. The spectral peaks were broad and not always centered very close to the original color. In view of the design of color film, the relatively broad absorption spectra of the dyes, and the vagaries of processing this should probably not have been so much of a surprise. Fortunately the spectral resolution appeared to be satisfactory for structured light imaging with a moderate number of band colors.

Band patterns with a variety of color sequences were produced, projected on test object and subjects, images of the resulting light patterns recorded on Kodak Ektar 1000 film, and these were then digitized on the film scanner. Surprisingly, the digitized bands were of unequal width in areas where they should have been of similar dimensions. (Figure 8.7)

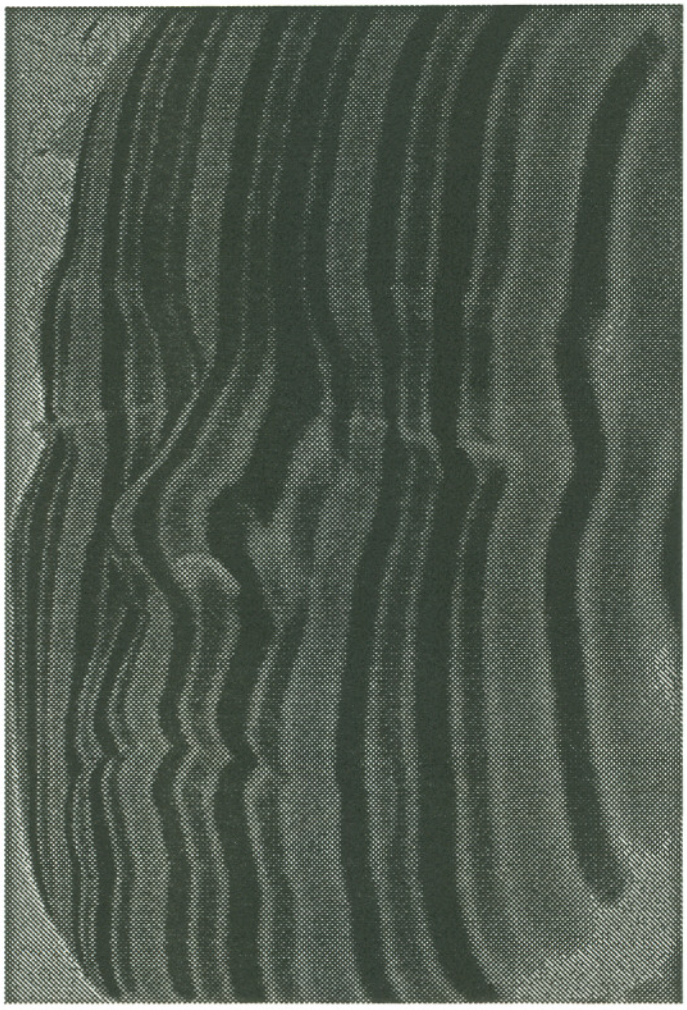


Figure 8.7 Image of face showing unequal width of projected color bands.

These differences in size were rather dramatic and clearly not apparent to an observer carefully examining the projected bands on the subject or object. These variations were probably a result of differential sensitivities of the three color layers in the film chosen for the tests. It was less evident in the original tests done with Kodachrome than with the Ektar, but the higher sensitivity of Ektar made the exposure times more practical. A system with a high intensity flash replacing the slide projector bulb would undoubtedly simplify these exposure issues.

A final series of tests was done using bands with two alternating colors, red and cyan, under the conditions described above. The bands on the digitized images were clear, sharp, and easily followed through eyebrows and across the lips. The discrepancy in the width of the bands noted before is also present in these images, but since the widening or narrowing of the bands should be symmetric around the center, there should be no significant loss of accuracy from assuming that the centers of the narrow bands represent isocontours for depths as described previously. (Figure 8.8)

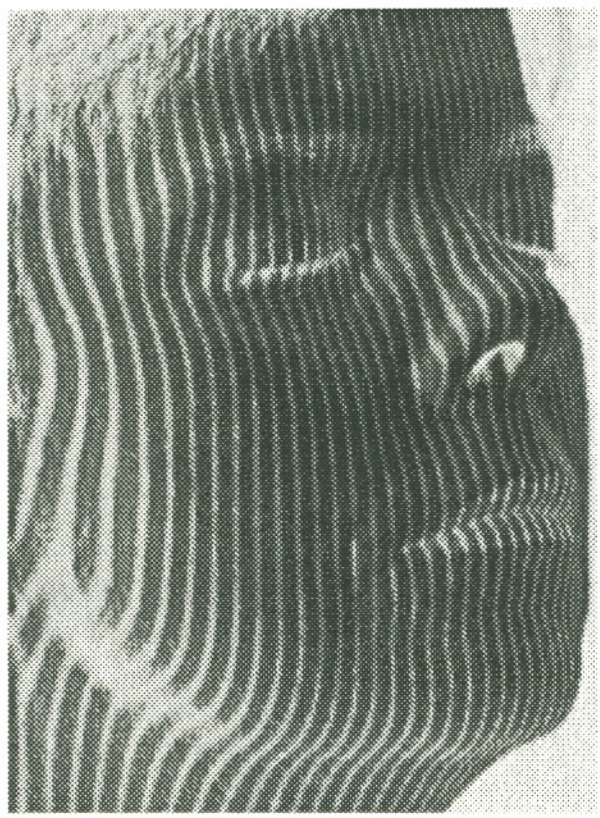


Figure 8.8 Image of face with bands from 2-color projection.

The color band patterns in 24 bit RGB format were enhanced using predominantly the color-separation capabilities in a commercial software package (Photostyler), the hair, clothing and other extraneous sections edited out of the image, converted to grey-scale images, (Figure 8.9)

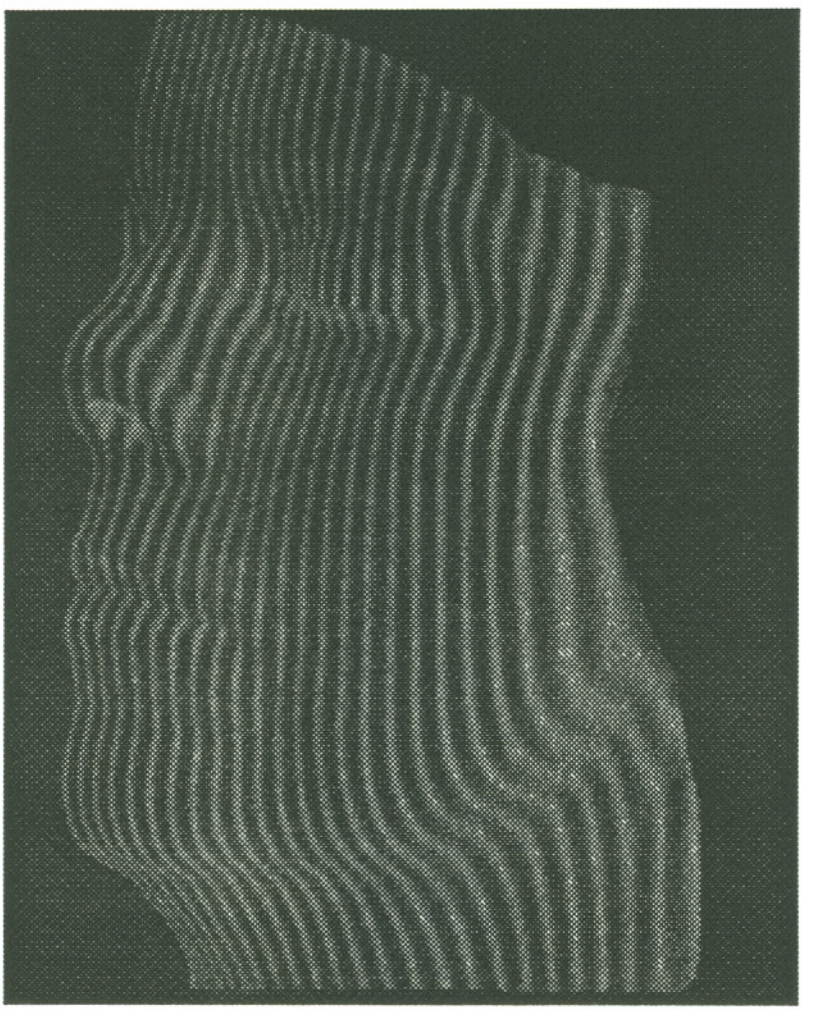


Figure 8.9 Enhancement of bands from 2-color projection. (Same subject as Fig. 8.8)

and the edges (boundaries between the original colors, now shades of grey) detected as described above using software written in C on a desktop computer. (Figure 8.10)

Figure 8.10 Edges extracted from Figure 8.9

The bands between the edges were labeled successively with values identifying their relative elevations such that the first band, representing the isocontour farthest from the camera, had a value of **x** and the subsequent bands **x + n*c**, where **n** is the order of the band and **c** is an arbitrary constant for demonstration purposes. No attempt was made to get absolute z coordinates for this demonstration but, as noted above, under controlled circumstances absolute z coordinates may be calculated from the band pattern when the geometry is known. In this case then the constant, **c**, would represent the geometrically-determined increment in the z coordinate between bands. (Figure 8.11)

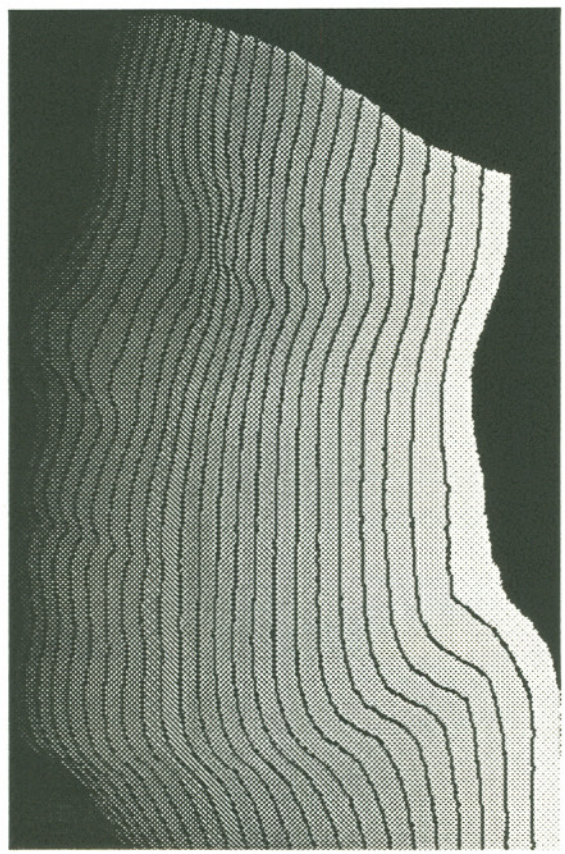


Figure 8.11 Isocontours intensity labeled by relative elevations

Another C program was then used to sample every 4th column and row in this file to create an array of xyz coordinates. This array was converted to a wire-frame type image using Mathcad. The resulting image, while crude because of the relatively low number of samples in the xyz file, clearly shows the desired z coordinate elevations. These contours are approximately 4 mm in width but the system should allow for 1-2 mm resolution without too much difficulty. (Figure 8.12)

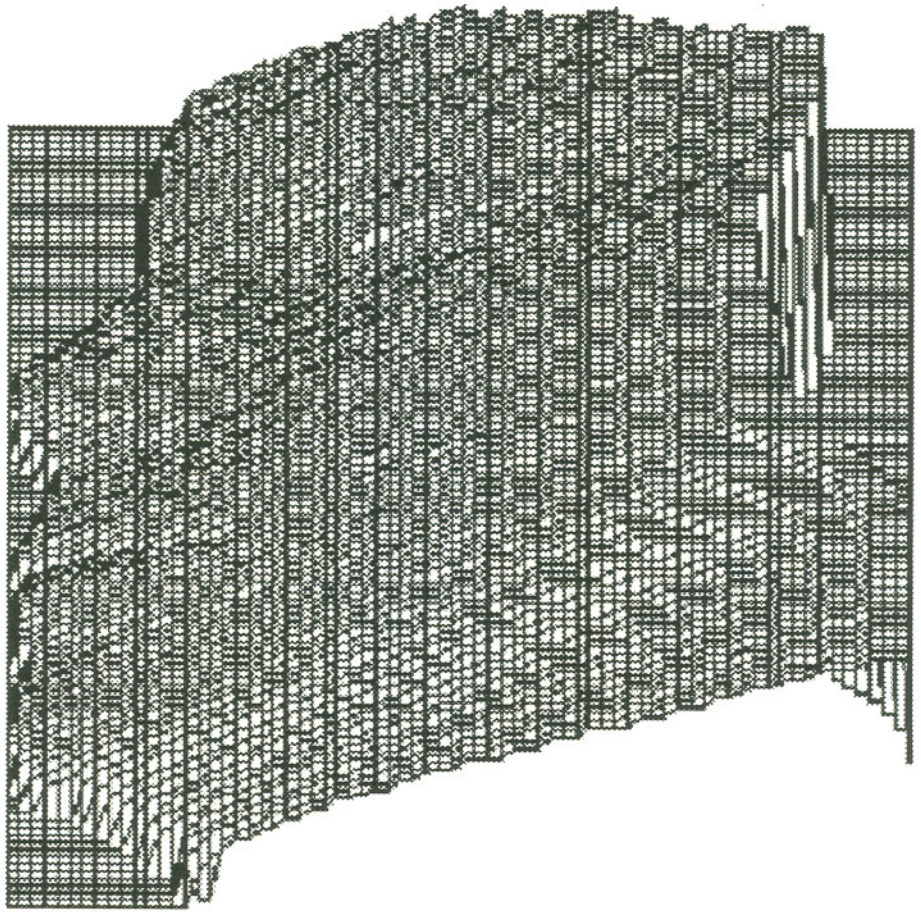


Figure 8.12 Surface elevations of face reconstructed from data from Fig 8.11

Multispectral structured light imaging of the human face appears to be a viable method to obtain three-dimensional coordinates of the surface. It will of course be necessary to image some test objects of known dimension in order to determine the accuracy of this approach. Since the geometry of the system precludes imaging the entire surface of the face at once, some way of combining multiple views into a single, registered surface must be devised. Two views could probably suffice and be obtained by a pair of cameras mounted similarly to the stereo apparatus described previously. Since the geometry would then be known, the reconstructed surfaces could presumably be assembled without too much difficulty.

If the patterns could be projected through filters with sharper absorption peaks and recorded with greater fidelity multispectral imaging would open up some interesting possibilities. If the entire spectrum were projected onto an object then the depth of that point is encoded by the color. For example, if the spectrum were projected with red closest to the camera and blue the farthest away and the width of the entire spectral band and distribution of colors within the band were known, then a point reflecting pure yellow would always be a known distance from the camera and similarly for other colors. In a video system capable of high spectral differentiation it might be feasible to do real-time 3-dimensional surface imaging since the calculation of the z coordinate would be trivial once the red, blue and green intensity values were known. The combination of multispectral imaging with structured light might also make stereometric analysis of the face or other smooth surfaces possible. The color encoding of the pattern would obviate the matching difficulties inherent with almost all other approaches.

CHAPTER 9

CONCLUSIONS

Advances in medical genetics and other clinical fields will continue to increase the demands for objective methodology to assess facial configuration. The complexity of the face makes this a very difficult challenge for clinicians and some form of imaging technology will undoubtedly be necessary to assist in the diagnostic process. Although 2-dimensional imaging may be useful for some purposes, the use of 3-dimensional surface reconstructions will likely be necessary for more sophisticated analyses.

Because of the limitations in the types of technologies which might be usable in clinical settings, obtaining accurate, high-resolution x, y, and z coordinate information for the surface of the face has proved to be much more challenging than initially appreciated by many people. The use of multispectral structured light appears to hold considerable promise for 3-dimensional imaging of the face and may be useful in some other settings because of the possibility of encoding depth in color. This could reduce the computational complexity for certain problems considerably and simplify real-time capture of depth information.

Once reliable surface imaging of the face is readily available, it will be necessary to devise appropriate analytical schemes to make optimal use of this wealth of information. The simplest approach would be to use the surface reconstructions to obtain point-to-point distance measurements or angular measurements - essentially a 3-dimensional analog of the photogrammetric methods described above. Somewhat more complex would be the application of Bookstein's method of invariants to this surface. More sophisticated approaches might include something along the line of subdivision of the surface reconstruction of the face into biologically interesting regions, developing a large database of normal and abnormal shapes for these different regions, then analyzing a face in question by comparing it region by region with the database and seeing how this

comparison fits against similar comparisons for known disorders. This approach should have a more intuitive feel to the clinician than a pure mathematical/statistical analysis or a computational approach such as the neural network described above for a 2-dimensional model.

REFERENCES

Bookstein FL: The Measurement of Biological Shape and Shape Change. Lecture Notes in Biomathematics - 24. Springer-Verlag, Berlin 1978.

Bookstein FL: The Geometry of Craniofacial Growth Invariants. American Journal of Orthodontics 83:221-234 1983.

Bracewell RN: The Hartley Transform. Oxford University Press, New York, 1986.

Clarren SK, Sampson PD, Larsen J, Donnell DJ, Barr HM, Bookstein FL, Martin DC, Streissguth AP: Facial Effects of Fetal Alcohol Exposure: Assessment by Photographs and Morphometric Analysis. Am J Med Genet 26: 651-666 1987.

Cohen PR, Feigenbaum EA: Handbook of Artificial Intelligence Volume 3. Addison-Wesley, Reading MA 1982

DiLiberti JH: Pattern Recognition of 2-Dimensional Biological Structures: The Use of Fourier Descriptors. David Smith Dysmorphology Workshop, Burlington, Vermont August 1986.

DiLiberti JH: D'Arcy Thompson Revisited: Practical Clinical Computerized Coordinate Transforms. Presentation at the David W Smith Dysmorphology Workshop. Lexington KY, August 1990.

DiLiberti JH, Olson DP: Photogrammetric Evaluation in Clinical Genetics: Theoretical Considerations and Experimental Results. American Journal of Medical Genetics 39:161-166 1991.

DiLiberti JH: Practical Clinical Coordinate Transformation in Dysmorphology. Journal of Medical Genetics, 1991 28:472-476.

DiLiberti JH, Pober B, et al.: A Comparison of Photogrammetric Measurements with Anthropometric Measurements: A Controlled Study. In preparation 1993.

Donahue BA, Bronzino JD, DiLiberti JH, Olson DP, Schweitzer LR, Walsh P: Application of a Neural Network in Recognizing Facial Expression. Proceedings of the 17th Annual Northeast Bioengineering Conference 1991.

Gonzalez RC, Woods RE: Digital Image Processing, Addison-Wesley, Reading MA, 1992.

Gorlin RJ, Cohen MM Jr., Levin LS : Syndromes of the Head and Neck. Oxford University Press, New York 1990.

Gould SJ: A Biological Homage to Mickey Mouse in The Panda's Thumb. WW Norton, New York 1980.

Grimson WEL: From Images to Surfaces. M.I.T. Press, Cambridge MA 1981.

Grunau RVE, Craig KD: Pain Expression in Neonates: Facial Action and Cry. Pain 28:395-410 1987.

Hall JG, Froster-Iskenius, Allanson JE: Handbook of Normal Physical Measurements. Oxford University Press, Oxford, 1989.

Hoehne KH, Delapaz RL, Bernstein R, Taylor RC: Combined Surface Display and Reformatting for the Three-Dimensional Analysis of Tomographic Data. Investigative Radiology 22:658-664 1987.

Hojo K, Umeda T, Yotsuya H, Yoshinobu I: Bull Tokyo Dent Coll 23:37-45 1982.

Jones KL: Smith's Recognizable Patterns of Human Malformation 4th Edition. WB Saunders, Philadelphia 1988.

Kodak Professional Black and White Films. Eastman Kodak Company, Rochester NY 1984.

Li J, Su XY, Guo LR: Improved Fourier Transform Profilometry for the Automatic Measurement of Three-dimensional Object Shapes. Optical Engineering 29:1439-1444 1990.

Moore KL: The Developing Human - Clinically Oriented Embryology. WB Saunders, Philadelphia 1973.

Pober B, DiLiberti JH, et al.: A Controlled, Multiexaminer Trial of Anthropometrics. In preparation 1993.

Press WH, Flannery BP, Teukolsky SA, Vettering WT: Numerical Recipes in C. The Art of Scientific Computing. Cambridge University Press. Cambridge 1988.

Rumelhart DE, McClelland JL: Parallel Distributed Processing Volume 1: M.I.T. Press, Cambridge MA, 1986.

Savara BS, Miller SH, Demuth RJ, Kawamoto HK: Biostereometrics and Computergraphics for Patients with Craniofacial Malformations: Diagnosis and Treatment Planning. Plastic and Reconstructive Surgery 75:495-499 1985.

Thompson D: On Growth and Form. Cambridge University Press, Cambridge 1961.

Winston PH: Artificial Intelligence 2nd Edition. Addison-Wesley, Reading MA 1984

Yakimovsky Y, Cunningham R: A System for Extracting Three-Dimensional Measurements from a Stereo Pair of TV Cameras. Computer Graphics and Image Processing 7:195-210 1978.

Young AW, Ellis HD (Editors): Handbook of Research on Face Processing. Elsevier Science Publishers 1989.

APPENDIX I

NOTE ON DIGITIZED IMAGES

Many of the illustrations were originally of fairly high resolution, either with a range of 256 shades of grey (8 bit) or full color (24 bit) images. They have been reproduced below as half-tone images using a 300 dot-per-inch laser printer. Obviously some details which were readily apparent in the original images have either been lost or greatly attenuated in visibility. Descriptions in the text may therefore not match the illustrations as accurately as if some other reproduction process could have been used.

APPENDIX II

TECHNICAL AND PROGRAMMING NOTES

All software written by the author was compiled using Microsoft C version 7.0, Microway NDP C version 1.4, or Computer Innovations C86 on Intel 386-33 and 486-50 based computers.

1. Page 14 Digitization and resampling of images.

The images used for this experiment were derived from a 35 mm transparency of a child with trisomy 21 (Down syndrome). The slide was illuminated using a device made by the author, which consisted of an old photographic enlarger head mounted horizontally on a wooden frame with a slide holder made out of wood and metal. On the opposite end of the platform a high-resolution black and white video camera was mounted on a spring-loaded frame which could be positioned for proper alignment between the camera and the slide. A 55 mm Micro-Nikkor lens was mounted on the video camera using a C-mount. The camera was attached to an Imaging Technologies capture board housed in an MS-DOS based, Intel 486-25 computer. The initial image was captured using the software provided with the image capture board, Image Pro, with a resolution of 512 x 512 pixels and 8 bits per pixel. The output file was in Tag Image File Format (TIFF). Subsequent manipulations were done using either TIFF files or raw data files without headers on occasion.

All processing of images 3.5 a - d was done using software written in C by the author. The original 512 x 512 image was sampled at the indicated lower resolutions and then the resulting files were re-expanded back to 512 x 512, using an 8 bit per pixel grey scale throughout. The re-expansion was done using bicubic interpolation in order to preserve smoothness of image intensities and avoid jagged, large square pixels. The bicubic interpolation code borrowed heavily from Numerical Recipes in C [Press et al 1988]. The resulting images were displayed on a high-resolution monitor using code

written by the author and then photographed and printed by the author on black and white paper using standard photographic darkroom procedures. The laser-printed images in this paper were printed using commercial software.

2. Page 28. Grid-warping.

The images used in this experiment were obtained as described in Note 1 above and the final output images were also produced as described in Note 1. All of the processing was done using C code written by the author, except for the section using bicubic interpolation as noted above. This program prompted the user to select a file to grid-warp, then read the file into memory and displayed it on the high-resolution (1024 x 768) screen. A cross-hair cursor, controlled by the operator, could then be moved to the location to be warped. This was assisted by numerical readout of the pixel locations on the side of the screen. Once this pixel was confirmed, the operator would then select the new location to which the first point would be stretched. The program then mapped the changed pixels using output to input mapping (to avoid gaps in the image) and bicubic interpolation. This process could be performed as many times as necessary in order to obtain the desired result. Further information on this project is contained in a recent publication [DiLiberti 1991].

3. Page 33. Neural network.

These experiments were conceived by the author and implemented by Barbara Donohue as her research project for a Master of Science degree in bioengineering at the Hartford Graduate Center. Most of the code was written in C on an Intel 486-25 computer by David Olson who worked in the author's lab at that time and who supervised Ms. Donohue with help from the author. Further information on this project is contained in a recent publication [Donohue et al. 1991]

4. Page 40 Stereometric imaging.

Images were captured using an apparatus similar to that described in Note 1. Initial attempts at z coordinate extraction were made using code implemented by the author and Lise Storc who worked in my lab at that time. All of the code for enhancing edges and details for stereo matching was written by the author. Subsequent attempts at stereo matching and all of the scan line analyses were done using C code written by the

author. The image filtering to demonstrate the lack of meaningful details in the scan lines was accomplished using Photostyler and the output from the scan line analyses was graphed using Harvard Graphics.

5. Page 53. Structured light - black and white.

The digitization of the black and white structured light images was accomplished as explained in Note 1. All of the filters and edge detection programs were written by the author, but the actual Hartley transform subroutine was from Bracewell [1986].

6. Page 57. Structured light - generation of colored bands.

The program to generate images of parallel bands with arbitrary widths, repeat patterns, and 24 bit color was written by the author as was the program to display the resulting files on the monitor for photography.

7. Page 59. Structured light - generation and capture of colored bands.

The band patterns were generated using a modification of the software written for the experiment in Note 6. They were transferred to the film recorder using Image Print software provided by the manufacturer of the film recorder. The facial images with projected band patterns were digitized using the Nikon scanner and software, Photostyler, provided by the manufacturer.

8. Page 64. Band enhancement and edge detection.

The programs/filters to enhance the color bands and detect and extract edges were written by the author using modifications of the code for black and white images described in Note 5. For convenience Photostyler was also used for some of these same (filtering and color separation) operations since the images could be manipulated and displayed in a well-integrated environment which tended to save time.

9. Page 65. Band labeling.

This was accomplished using Photostyler for the reasons noted above in Note 8.

10. Page 67. Sampling of labeled images.

Sampling was done at regular intervals using code written by the author. The image file to be sampled was in TIFF format and the sampling was done in a way to create an output file compatible in format and array size with Mathcad Version 3. This limited the output array size to fewer than 8,000 points, hence the relatively coarse resolution in Figure 8.12.

BIOGRAPHICAL SKETCH

John DiLiberti was born on July 23, 1945 in Jersey City, New Jersey. He attended Princeton University receiving Bachelor of Arts degree *Cum laude* in Chemistry in 1966 and subsequently the University of Cincinnati College of Medicine, receiving the Doctor of Medicine degree in 1970. Post-doctoral training in Pediatrics, Genetics, and Metabolism was obtained at Temple University / St. Christopher's Hospital for Children in Philadelphia. His undergraduate thesis research involved computerized modelling of molecular orbital energy levels in transition metal complexes. Later areas of special interest include the applications of computers in medical care, especially in human genetics, as well as clinical research into genetic disorders.

From 1975 until 1988, he was on the faculty of Oregon Health Sciences University as Assistant and then Associate Professor of Pediatrics. From 1988 until 1991 he was Professor and Associate Chairman of the Department of Pediatrics at the University of Connecticut School of Medicine, and 1991 through 1993 Professor of Pediatrics at Northeastern Ohio Universities College of Medicine.

Publications:

1. Gruskin A, DiLiberti JH et al:
   Calcium Excretion in Cystic Fibrosis. J Ped 82:344 1973. (Clinical memorandum)

2. DeForest A, Parker P, DiLiberti JH et al:
   The effect of Breast-feeding on the Antibody Response of Infants to Trivalent Oral Poliovirus Vaccine. J Ped 83:93 1973.

3. DiLiberti JH, O'Brien ML, Turner T:
   The Use of Physostigmine as an Antidote in Accidental Diazepam Intoxication
   J Ped 86:106 1975.

4. DiLiberti JH and O'Brien ML:
Pseudotumor Cerebri Following Patent Ductus Arteriosus Ligation. J Ped 87:489 1975. Clinical Memorandum)

5. DiLiberti JH, McKean R, Hecht F:
Progressive Tenosynovitis with Systemic Involvement - A New Heritable Disorder of Connective Tissue? Birth Defects, Original Article Series 11:81 1975.

6. St Clair EH, DiLiberti JH, O'Brien ML:
Observations of an Infant Born to a Mother with Botulism. J Ped 87:658 1975. (Clinical Memorandum)

7. DiLiberti JH, McKean R, Webb MJ, Williams G:
Trisomy 5p: Delineation of Clinical Features. Birth Defects, Original Article Series 13:185-194 1977.

8. DiLiberti JH, Bartel SJ, Humphry TR, Pang A:
Acute Monoarticular Arthritis in Association with Varicella. Clin Ped 16:663-4 1977.

9. Levy JS, Roy S, Hutcheson RH, Gruskin AB, Hellerstein S, Linshaw M, Alexander S, DiLiberti JH et al:
Infant Metabolic Alkalosis and Soy-Based Formula. Morbidity and Mortality Weekly Report 28:30 1979.

10. Illingworth DR, Connor WE, Lin DS, DiLiberti JH:
Lipid Metabolism in Abetalipoproteinemia: A Study of Cholesterol Absorption and Sterol Balance in Two Patients. Gastroenterology 78:68-75 1980.

11. Linshaw M, Harrison H, Gruskin A, Prebis J, Harris J, Stein R, Jayaram MR, Preston D, DiLiberti JH et al:
Hypochloremic Alkalosis in Infants Associated with Soy Protein Formula. J Ped 96:635-640 1980.

12. DiLiberti JH:
Familial Omphalocele: Analysis of Risk Factors and a Case Report. Amer J Med Genet 13:263-268 1982.

13. DiLiberti JH and Tarlow S:
Bone and Joint Complications of Haemophilus Influenzae Meningitis. Clin Ped 103:7-10 1982.

14. Finnell RH and DiLiberti JH:
Hydantoin-induced Teratogenesis: are Arene Oxide Intermediates Really Responsible? Helvetica Pediatrica Acta 38:171-177 1983.

15. DiLiberti JH, Budden S, Schimschock J:
The Ruvalcaba-Myhre-Smith Syndrome: a Case with Probable Autosomal Dominant Inheritance and Additional Manifestations. Amer J Med Genet 15:491-496 1983.

16. DiLiberti JH, D'Agostino AN, Cole G:
Needle Muscle Biopsy in Infants and Children. J Ped 103:566-570 1983.

17. DiLiberti JH, D'Agostino AN, Ruvalcaba RHA, Schimschock JR:
A New Lipid Storage Myopathy in the Ruvalcaba-Myhre-Smith Syndrome. Amer J Med Genet 18:163-167 1984.

18. DiLiberti JH and Laxer KD:
Abnormalities in Brain Stem Auditory Evoked Potentials in Reye's Syndrome. Clin Ped 23:238-239 1984.

19. DiLiberti JH, McMurry M, Connor WE, Alaupovic P:
Hypercholesterolemia Associated with Alpha-1 Antitrypsin Deficiency and Hepatitis: Lipoprotein and Apoprotein Determinations, Sterol Balance and Treatment. Amer J Med Sci 288:81-85 1984.

20. DiLiberti JH, Farndon PA, Dennis NR, Curry CJR:
The Fetal Valproate Syndrome. Am J Med Genet 19:473-482 1984

21. Shapiro SD, Hansen KL, Pasztor L, DiLiberti JH, et al.:
Deletions of the Long Arm of Chromosome 10. Am J Med Genet 1985.

22. Rogers MF, Rutherford GW, Alexander SA, DiLiberti JH:
A Population Based Study of the Hemolytic Uremic Syndrome In Oregon, 1979 - 1982. Am J Epidemiology 123:137-142 1986.

23. Landrigan PJ, DiLiberti JH, Graef JW, Jackson RJ, Nathanson G:
Involuntary Smoking - A Hazard to Children. Pediatrics 77:755-57 1986.

24. Landrigan PJ, DiLiberti JH, Graef JW, Jackson RJ, Nathanson G:
Asbestos Exposure in Schools. Pediatrics 79:301-305 1987.

25. Landrigan PJ, DiLiberti JH, Gehlbach SH, et al:
Statement on Childhood Lead Poisoning. Pediatrics 79:457-65 1987.

26. Gretzula JC, Hevia O, Schachner LS, DiLiberti JH, et al.:
Ruvalcaba-Myhre-Smith Syndrome. Pediatric Dermatology. 5:28-32 1988.

27. DiLiberti JH:
Use of Computers in Dysmorphology. J Medical Genetics. 25:445-453 1988.

28. Daya MR, Burton BT, Schleiss MR, DiLiberti JH:
Recurrent Seizures Following Mucosal Application of TAC. Annals of Emergency Medicine. 17(6):646-648 1988.

29. Rosenberg NL, DiLiberti JH, Buist NRM, Andrews T:
Myotonic Dystrophy and Hyperparathyroidism: Association with Neurofibromatosis, and Multiple Endocrine Adenomatosis Type 2A. J Neurol, Neurosurg, Psychiatry. 51:1578-1580 1988.

30. Jackson RJ, DiLiberti JH, Landrigan PJ, Nathenson G, Needleman HL et al.:
Radon Exposure: A Hazard to Children. Pediatrics. 83:799-802 1989.

31. Murphy M, Greenberg F, Wilson G, Hughes M, DiLiberti JH:
Williams Syndrome in Twins. American Journal of Medical Genetics (Supplement) 6:97-99 1990.

32. DiLiberti JH, Olson DP:
Photogrammetric Evaluation in Clinical Genetics: Theoretical Considerations and Experimental Results. Am J Med Genet 39:161-166 1991.

33. DiLiberti JH:
Practical Clinical Coordinate Transformation in Dysmorphology. Journal of Medical Genetics, 1991 28:472-476

34. Donahue BA, Bronzino JD, DiLiberti JH, Olson DP, Schweitzer LR, Walsh P:
Application of a Neural Network in Recognizing Facial Expression. Proceedings of the 17th Annual Northeast Bioengineering Conference 1991.

35. DiLiberti JH:
Lipid Storage Myopathy and Macrocephaly - Clinical and Histochemical Correlations. Journal of Medical Genetics 29:46-49 1992.

Book Chapters:

1. Zonana J, DiLiberti JH:
Congenital Renal Disease. in Emery A and Rimoin D (eds) Principles and Practice of Medical Genetics. Livingstone Churchill, Edinburgh 1983, pp 987-1002.

2. DiLiberti JH:
Shwachman Syndrome. in Buyse ML (ed) Birth Defects Encyclopedia. Blackwell Scientific Publications 1990.

3. DiLiberti JH:
Overgrowth: Ruvalcaba-Myhre-Smith Type in Buyse ML (ed) Birth Defects Encyclopedia. Blackwell Scientific Publications 1990.

4. DiLiberti JH:
Familial Arthritis - Arteritis Syndrome in Buyse ML (ed) Birth Defects Encyclopedia. Blackwell Scientific Publications 1990.

5. Zonana J, DiLiberti JH:
Congenital Renal Disease. in Emery A and Rimoin D (eds) Principles and Practice of Medical Genetics. 2nd Edition Livingstone Churchill, Edinburgh 1990.