

# Face recognition based on depth maps and surface curvature

Gaile G. Gordon  
Harvard Robotics Laboratory  
Pierce Hall G12b  
Harvard University  
Cambridge, MA 02138

## Abstract

This paper explores the representation of the human face by features based on shape and curvature of the face surface. Curvature captures many features necessary to accurately describe the face, such as the shape of the forehead, jaw line, and cheeks, which are not easily detected from standard intensity images. Moreover, the value of curvature at a point on the surface is also viewpoint invariant. Until recently range data of high enough resolution and accuracy to perform useful curvature calculations on the scale of the human face had been unavailable. Although several researchers have worked on the problem of interpreting range data from curved (although usually highly geometrically structured) surfaces, the main approaches have centered on segmentation by signs of mean and Gaussian curvature which have not proved sufficient for classification of human faces. This paper details the calculation of principal curvature for our particular data set, the calculation of general surface descriptors based on curvature, and the calculation of face specific descriptors based both on curvature features and *a priori* knowledge about the structure of the face. These face specific descriptors can be incorporated into many different recognition strategies. We describe a system which implements one such strategy, depth template comparison, giving excellent recognition rates in our test cases.

## 1 Introduction

The face recognition problem has received a great deal of attention in the computer vision community. In some senses it is one of the most difficult of all the visual recognition tasks because it requires differentiating among objects which vary only subtly from each other. Yet humans perform the task effortlessly many times a day and have the capacity for storing and distinguishing among thousands of faces. Very little progress has been made in developing a robust automatic face classification system. It is quite possible that the reason for this lack of progress is impoverished input data. The human sensing system clearly provides a very rich set of information, yet we attempt to solve some of the same sensing tasks from single eight bit grey scale images. We propose that some of the missing information in the case of the face recognition problem involves the shape and curvature of the face. Range sensors now produce data of such resolution and accuracy that, with careful treatment of noise, accurate curvature calculations can be made for the human face. The availability of curvature data opens up many new avenues for face recognition methods. Curvature, as a local surface property of the object, also has the valuable characteristic of being viewpoint invariant. In the following sections we introduce the curvature calculation method, and a variety of both general and face specific descriptors based on depth and curvature. We demonstrate the use of these descriptors in the implementation of a flexible template based face recognition system.

## 1.1 Background

Most face recognition research has centered on feature extraction from intensity images. In [Fischler and Elschlager 1973] faces were used as an example to introduce the “Linear Embedding Algorithm”. This method finds the placement of pictorial structures relative to each other by minimizing a cost functional based on a spring model. The features themselves, however, were not described in any respect except position, and the system did not attempt to distinguish between faces. More recently in [Yuille, Cohen, and Hallinan 1989], flexible templates were used to describe the eyes and the mouth with a small set of parameters. Since these templates have not yet been incorporated into a full face recognition system, we must wait to evaluate their eventual usefulness. Perhaps the only fully automated face recognition system tested to date is Kanade’s [Kanade 1973] which used 13 parameters to describe and match 40 frontal view face images made up of 2 views of 20 different faces. The parameters described spatial relationships among several feature points detected in the intensity image. The basic features were fairly simplistic, including the extent of the eye and mouth regions, location of nostrils, and approximation to the symmetry line of the face. No compensation was made for variation of head position, background, or unusual lighting conditions. Despite the small number of test cases and the strong restrictions on the input images, the success rate of 15 correct matches from the 20 faces makes it one of the most successful face recognition systems!

Other intensity based approaches to face recognition are based on general pattern recognition techniques and make no use of face specific knowledge. These systems often show very good recognition rates over a restricted set of inputs. However, they are not very flexible with respect to orientation or scale changes for instance, and therefore require storage of multiple “templates” to perform recognition over varying conditions. This makes them impractical for systems which must deal with large numbers of subjects; but for other types of applications these methods can be very useful [Turk, Pentland 1991].

One clear difficulty in these approaches is that features are extracted using image contrast. Thus many features of the face are difficult to detect or measure because of variability of lighting conditions or low dynamic range in the input image. Even the highest contrast features of the face, such as the eyes, are a challenge to identify and describe reliably [Hallinan 1990]. Low contrast features such as shape of jaw boundary, cheeks, and forehead are currently impossible to describe from general intensity images.

Even if we could extract reliably and accurately the position and descriptions of the standard facial features (eyes, nose, mouth, outline of face area), there is good deal of evidence in the psychology literature to suggest that this is not sufficient for humans to perform individuation among many faces. For example, Rhodes, Brennan, and Carey [1987] have demonstrated that we are not very good at identifying faces from simple line drawings. Similarly Davies, Ellis, and Shepherd [1978] showed that famous faces were recognized more accurately from photographs than from line outlines; subjects performed at a 90% accuracy level with photographs, and at a 23% accuracy level with line outlines. These experiments suggest that it may also be necessary to describe the shape of the face, exactly these areas of the face which are difficult to describe with standard intensity based methods. Although it is unlikely that humans base their representation or comparison of shape on the accurate perception of depth, we propose the use of depth data because at our current state of technology it is the most straight forward way to input or record complex shape information for machine analysis. It is likely that lessons learned about critical stages in machine face recognition using depth data will offer insight into the way humans recognize faces.

In his experiments on man-machine interaction in face recognition, Harmon presents the sole example of a system which explicitly incorporates shape information into the task of face recognition [Harmon 1979]. In his system human subjects identified qualitative facial features from photographs. Each face was encoded by a set of 21 feature descriptors. The set of feature descriptors he used included forehead, cheeks, and chin. Any given face was encoded by assigning a value from 1 to 5 to each feature. For the forehead and chin these values corresponded to the range from “receding” to “bulging”, and for the cheeks to the range from “sunken” to “full”. This technique showed a relatively good success rate; 99% of the time the database of 255 faces could be reduced to ten which included the stimulus face.

## 1.2 The Use of Depth Imagery

Advances in range finding technology have made it possible to trace the shape of the face at a resolution of 0.4 mm or smaller, allowing accurate calculation of surface curvatures. Segmentation and interpretation of general range images using surface curvatures has been previously studied by many researchers ([Besl and Jain 1986], [Fan, Medioni, and Nevatia 1985], [Vemuri, Mitiche, and Aggarwal 1986]). A good overview of this literature is found in [Besl and Jain 1985, and 1986].

The bulk of this literature focuses on recognition of geometrically simple machined objects, attempting to classify surfaces into planar regions, spherical regions, or surfaces of revolution. The most common approach to region segmentation is based on signs of Gaussian and mean curvature; this allows four surface types: convex, concave, and two types of saddle regions. Besl and Jain [1986] also considered regions of zero mean or Gaussian curvature; this segmentation produces a total of eight visible surface types. Fan, Medioni, and Nevatia [1985] use sign of mean and Gaussian curvature but also consider the local maxima in maximum curvature. Vemuri, Mitiche, and Aggarwal [1986] look only at regions marked by the sign of the Gaussian curvature, and spherical and planar umbilic regions. Umbilic regions are those regions in which the two principal curvatures are equal (or in practical terms,  $|k_{max} - k_{min}| <$  threshold). Segmentation is followed by model based analysis. These type of methods are not sufficient to classify with more natural, smoothly curving objects. Our surface categorization must contain information about not just curvature signs, but also principal curvature magnitude and direction, umbilic points, and extrema in both principle curvatures. We will also need to incorporate object specific knowledge into the segmentation process itself.

NTT's Human Interface Lab in Japan has also examined performing face recognition from range data. In a study [Masui, Akamatsu, Suenaga 1990] using data from 20 faces, and *manual* feature detection for normalization of view, they claim recognition rates of 94% to 99% using only the most basic match criterion of minimizing the absolute value of distance at each point between two normalized views of the face. Their viewing conditions are highly restricted, however, since they do not perform any feature detection. For instance, they make depth comparisons over the whole field of the image (head). This method has the obvious drawback that large differences in superficial detail such as hair or collar contours will outweigh subtle contributions from the face area. In current experiments the subjects wear bathing caps to minimize these types of errors. Shape and dimensions of the ears will also contribute highly to recognition; however, this information is often not available under normal viewing conditions. They also calculate the depth differences between points by using correspondence on a cartesian grid; therefore a difference in the widths of the two faces will weigh more heavily in the comparison than for instance differences in cheek or nose shape. This effect may not be desirable. Despite these problems their results are very encouraging.

## 2 Calculation of Curvature Maps

The data used in this paper were generated by a rotating laser scanner system. Depth is stored in a cylindrical coordinate system  $f(\theta, y)$ , at each point measuring distance from the central vertical axis to the surface for the given  $\theta$  and  $y$  coordinates. Each data set has 256 equally spaced samples along the vertical axis,  $y$ , and 512 equally spaced samples in  $\theta$  from 0 to  $2\pi$ . Figure 1(a) shows an example depth file in parameter space. Figure 1(b) shows a view of the same data set rendered from a polygon model made by connecting nearest neighbors in a subsampled depth map. This style of representing range data has many useful properties. Importantly it gives a panoramic perspective without complex reconstruction techniques. A rotation of the object about the  $y$  axis maps to a translation in  $\theta$  in the parameter space. A translation of the object within the range of the sensor will map to a dilation in parameter space. The mapping from parameter space to absolute 3D coordinates is straightforward.

This representation of the range data gives us a surface which is embedded in a three dimensional space, parameterized in two dimensions. Since the parameterization is different from the more traditional range case which gives depth as a function  $f(x, y)$ , the formulation of curvature calculations must reflect these differences. Let us briefly review the surface curvature measures and their calculation in our case. For a more detailed background in differential geometry consult [Millman and Parker 1977].

At each point  $P$  on the surface a curve is formed by the intersection of the surface and the normal plane in a given tangent direction  $\vec{t}_i$ . The curvature of this planar curve is the normal curvature  $\kappa_n$  at  $P$  in the direction  $\vec{t}_i$ . The maximum and minimum normal curvatures at a point define the *principal curvatures*,  $\kappa_{max}$  and  $\kappa_{min}$ ; the



Figure 1: (a) Depth of face parameterized as  $f(\theta, y)$  (Leonard Nimoy as Spock), (b) rendered polygonal model of face composed from coarse sampling of depth data.

directions  $\vec{t}_{max}$  and  $\vec{t}_{min}$  associated with each principal curvature are the *principal directions* of the surface at  $P$  and by Euler's theorem are orthogonal.

The principal curvatures and the principal directions are given by the eigenvalues and eigenvectors of the shape operator matrix:

$$L = DG^{-1}, \quad (1)$$

where  $D$  is the second fundamental form of the surface and  $G$  is the first fundamental form.

We must calculate the fundamental forms for the surface in our cylindrical parameterization [Gordon 1990]. For a point  $P = f(\theta, y)$ , the first fundamental form is given by:

$$G = \begin{bmatrix} f_\theta^2 + f^2 & f_y f_\theta \\ f_y f_\theta & f_y^2 + 1 \end{bmatrix}, \quad (2)$$

where subscripts indicate partial differentiation. The second fundamental form is given by:

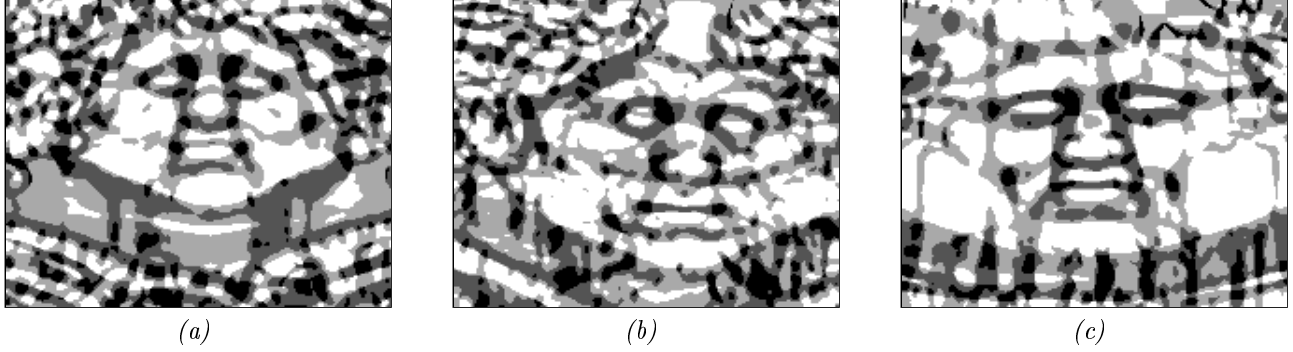
$$D = \frac{1}{f_\theta^2 + f^2(1 + f_y^2)} \begin{bmatrix} f_{\theta\theta}f - 2f_\theta^2 - f^2 & f_{y\theta} - f_\theta f_y \\ f_{y\theta} - f_\theta f_y & f f_{yy} \end{bmatrix}. \quad (3)$$

The *Gaussian curvature*,  $K$ , at a point is defined as the product  $\kappa_{max}\kappa_{min}$ , and is also the determinant of the shape operator matrix,  $|L|$ . The *mean curvature*,  $H$ , is  $\kappa_{max} + \kappa_{min}/2$ , and therefore can also be found as half the trace of  $L$ .

## 2.1 Smoothing Considerations

In practice, because these curvature calculations contain second partial derivatives, they are extremely sensitive to noise in the data. In many cases a great deal of smoothing is required before any stability can be achieved in curvature measurements. We would like to smooth the surface as little as possible; over smoothing will actually modify the surface phenomena we are trying to measure. As our intuition should indicate, this lower bound on smoothing level is a function of the noise level in the data and the curvature of the surface. It can be argued [Gordon 1990] that contribution of noise in the measurement of curvature is related to  $\Delta(n_i * G(\sigma_s))$ , where  $\Delta$  is the Laplacian operator,  $n_i$  is noise which is modeled as a Gaussian process with zero mean and standard deviation  $\sigma_n$ , and  $G(\sigma_s)$  is a Gaussian smoothing filter. We approximate the standard deviation of this error contribution by

$$\sigma_{error} = SD(\Delta(n_i * G(\sigma_s))) \approx \frac{\sigma_n}{\epsilon^2} \sqrt{\frac{3}{8\sqrt{\pi}\sigma_s^5}}, \quad (4)$$



*Figure 2 : Segmentation of three different faces by sign of Gaussian and mean curvature: concave (black), convex (white), and two types of saddle, positive mean curvature (light grey), and negative mean curvature (dark grey). Data is shown in  $(\theta, y)$  space.*

where  $\epsilon$  is the interpixel spacing in the sampled data.

For stable curvature measurements we would like the error due to noise to be at least an order of magnitude less than the curvature values we would like to measure. Thus, we need to compare the value of this function,  $\sigma_{error}$ , with the range of curvatures on the surface of the face. We used data taken from a smooth machined cylinder of known radius to estimate the noise level in our data. It can be approximated by a Gaussian with standard deviation  $\sigma_n = .15mm$ . An approximate range of curvatures on the face is from  $.23mm^{-1}$ , at the bridge of the nose, to  $.018mm^{-1}$ , the average curvature of the head itself. If we use an average interpixel spacing for our data,  $1.2mm$ , we find the lower bound on smoothing in the nose region is  $\sigma_s = 1.5$ , whereas for the cylinder data it is  $\sigma_s = 3.75$ . There are few areas of discontinuous depth on the surface of the face; the edge of the nostril usually being the only example. These areas should be considered as having infinite curvature, and therefore require no smoothing.

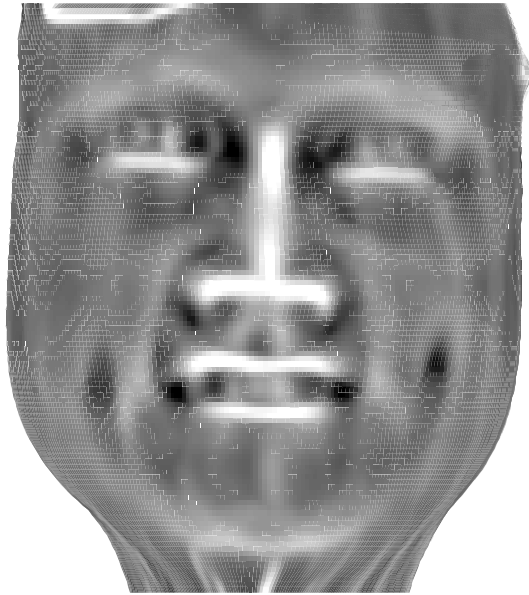
If we simply smoothed the whole image with the highest level of smoothing required we would severely modify the value and location of curvature features in high curvature areas (more discussion of types of smoothing and their effect on curvature measurement is given in [Gordon 1989]). Instead for accurate and reliable measurements we must use an adaptive smoothing method based on local estimates of curvature (not slope!). These local estimates of curvature can be based on the data, or on both the data and prior knowledge of the surface.

In this implementation we precompute curvature values using several different levels of smoothing. We use the curvature maps computed with a relatively low smoothing level to establish the location of features based on curvature extrema, and thus the general orientation of the face. We then use prior knowledge of the face structure and its expected local curvature to select curvature values from the precomputed set with the most locally appropriate smoothing level.

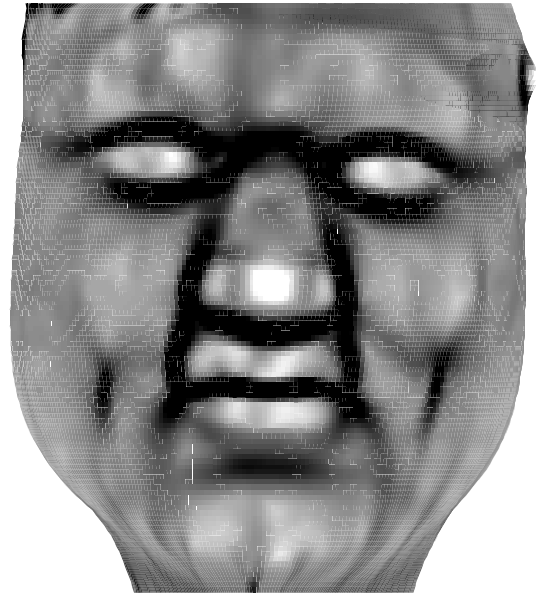
## 2.2 Results of principle curvature calculations

We can now use these methods to compute principal curvatures on the surface. An example of principle curvature maps calculated for one of the faces in the data base is given in figure 3. We have mapped the curvature maps to cartesian coordinates, and represent the magnitude of curvature as intensity. Darker values are lower than light values. The vector fields comprising the principal curvature directions are shown for a subsampled lattice as projected onto the plane  $z = 0$ . The umbilic points are also marked, their computation will be discussed below.

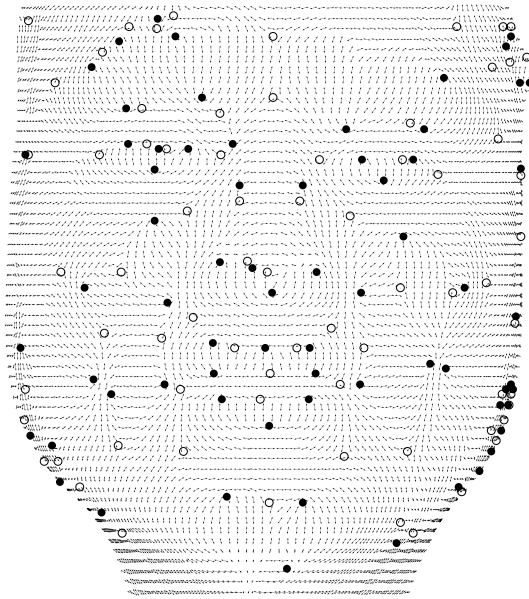
Segmentation by sign of Gaussian and mean curvature is very straightforward. Four kinds of regions are produced:  $K+,H+$  are convex,  $K+,H-$  are concave,  $K-,H+$  is a saddle with  $k_{max} + k_{min} > 0$ , and  $K-,H-$  is a saddle with  $k_{max} + k_{min} < 0$ . Figure 2 shows the resulting regions for three different sample faces.



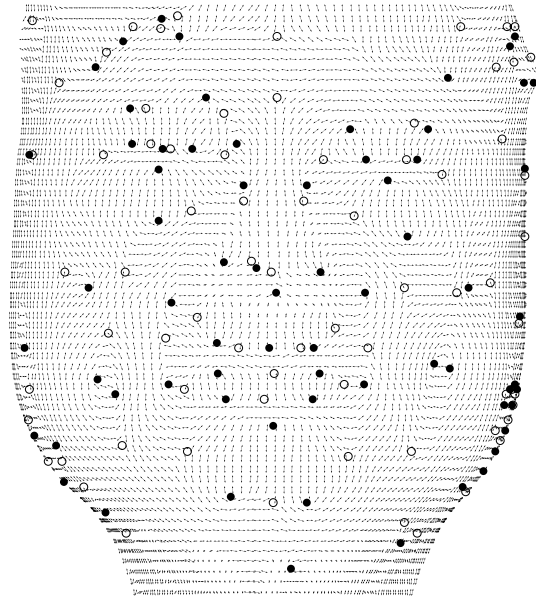
(a)



(b)



(c)

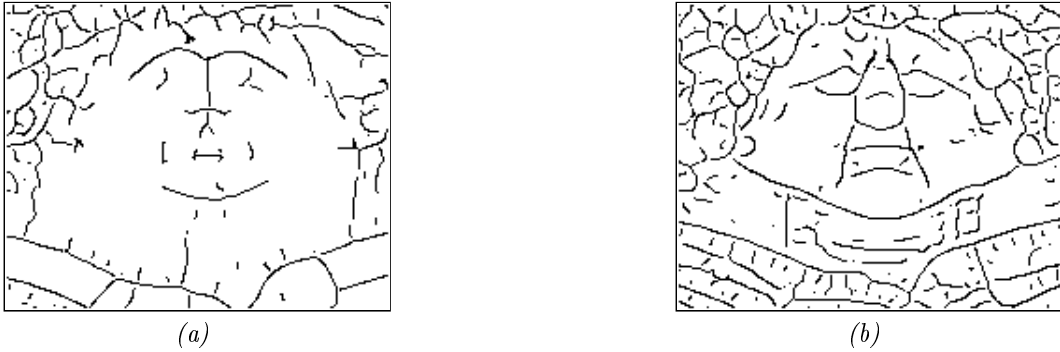


(d)

*Figure 3: Principle curvatures for a single face: magnitude (a) and direction (c) of maximum curvature, magnitude (b) and direction (d) of minimum curvature. Umbilic points are marked in (c) & (d); filled circles are points with positive index and open circles are points with negative index.*

### 2.3 Calculation of other general surface descriptors

Let us first consider the computation of extrema in  $\kappa_{max}$  and  $\kappa_{min}$ . Simply thresholding the maximum and minimum curvature maps yields large regions of extreme curvature; what we are really interested in, however, are ridge and valley lines. We define ridge lines to be local maxima in  $\kappa_{max}$  along the line of maximum curvature and valley lines similarly to be local minima in  $\kappa_{min}$  along the line of minimum curvature. Therefore we can suppress all points in the thresholded regions which are not local extrema in the direction indicated by the appropriate principle curvature vector. We use a bilinear interpolation among the local neighbors of the pixel to evaluate whether a point represents a local extrema. It is important to note that setting the thresholds in this case is not the sensitive process it is in edge detection. These thresholds are absolute measurements related to the physical structure of the surface. Figure 4 shows ridge and valley lines for one of the sample faces. Note how clearly the characteristic features of the face are displayed by these extrema.



*Figure 4 : (a) Ridge lines: local maxima of ( $k_{max} > thresh_r$ ), and (b) valley lines: local minima of ( $k_{min} < thresh_r$ ). Shown for the same face.*

Natural objects, such as faces, tend not to have large planar regions or spherical regions, but instead have isolated umbilic *points* which mark singularities in the lines of principal curvature. In practice it is difficult to locate umbilic points simply by comparing the two principle curvatures at a point. The curvatures will rarely be exactly equal; there will be errors due to noise, finite precision, grid interpolation. There is no way to automatically set a threshold such that only the point closest to the actual umbilic point is detected. We can instead consider the umbilic points in their role as singularities of the principal direction fields. A nonzero *index* of the direction field around a point is a more reliable indicator of an umbilic point since it takes into account the neighborhood around the point instead of only the curvature values at a single point [Sanders, Zucker 1991]. The index of a vector field can be described intuitively as the number of counterclockwise rotations we see in the vector field as we travel counterclockwise around the point. The index at umbilic points will be either  $1/2$  or  $-1/2$ . These two different types of umbilic points give rise to distinct patterns in the lines of curvature. In one the lines bifurcate as they approach the point, and in the other they terminate at the umbilic point. See figure 3c,d.

As these examples make clear, curvature descriptions of the human face are rich in information. Like finger prints they are as unique as the individual. We have at our disposal many tools with which to dissect and quantify this information, which is the subject of the next section.

### 3 Application to face recognition

A common denominator of many “face recognition” tasks is the need to represent the face quantitatively in a form which can then be used to compare one face to another. Such a representation can be used to match a given face to those stored in a data base. Or, in verification applications, we may wish to test whether the presenter of a pass key, such as a bank card, is actually the person who owns it. This can be accomplished without a data base by computing a representation of the person’s face and comparing it to one stored on the pass key.

With such a rich set of descriptors available there are many ways we can go about structuring a comparison strategy. We can categorize potential methods into two general approaches:

- template comparisons directly from depth or curvature values or,
- comparisons in feature space, where the feature values are descriptors (eg. distance between eyes, maximum curvature in cheek region) computed from the depth or curvature values.

These categories can also be applied to intensity based face recognition methods: the work of Turk and Pentland would fit into the first category, whereas the second category is typified by Kanade and the deformable templates of Yuille, Cohen, and Hallinan. There is actually some overlap in the implementation of the two categories. In order to make valid comparisons directly between depth maps of different faces (the first approach) we will need to use limited feature detection (the second approach) to register the images and compute effective template boundaries. Because of its basic nature and the implementation overlap, the first approach is a good place to begin investigation of face comparison from depth data.

More specifically, we wish to compare the surfaces of two similar objects by first normalizing the position of both surfaces with respect to a small set of detectable common feature points, and then calculating the volume of the difference area between them. In the case of faces, we hope that the difference between two instances of the same face will be less than that between two different faces. One obvious shortcoming of this method is that it doesn't allow for elastic changes in the surface such as varying facial expression. For instance, large differences will be found between two face images if the mouth is open in one and closed in the other, even if the two images represent different views of the same person. To reduce these effects we can define regions of the face over which comparisons are most meaningful. For instance the mouth and chin region are most highly affected by expression change, so we may not wish to consider differences in those regions. Similarly it is important that we be able to differentiate between skin regions and hair regions; not only is the data unreliable and noisy in the hair, but differences in hair style can contribute high values to the difference volume which would outweigh the more subtle but critical differences in the face itself.

The following discussion will present the implementation of the feature extraction and template matching process as well as experimental results.

### 3.1 Calculation of face specific features

The most unique aspect of the face in terms of range data is the nose. If the nose can be located reliably, it can be used to put constraints on the location of eyes and other features. There are at least three properties useful in locating the nose. Most obviously, it sticks out from the rest of the face. It also has a characteristic roof-like shape, and falls approximately on the symmetry plane of the face. In the current implementation we make use of the first two properties. The first step is to search for instances of roughly vertical connected ridges in the depth map. This is done by marking points of positive to negative sign change in  $d(radius)/d\theta$  for which an additional threshold constraint on  $|d(radius)/d\theta|$  on both sides of the nose is met. These requirements are local to the several pixels along the top of the ridge. Connected ridges over a trivial size are kept as potential nose ridge sites.

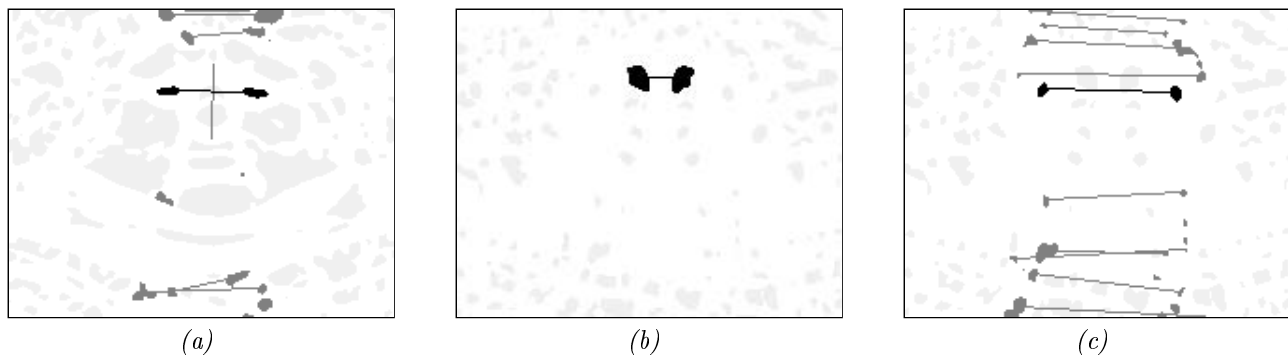
The second stage is to look for areas of high average radius surrounded on 3 sides by relative low average radius. This test detects protrusions in the depth data. It is performed by convolution with a modified center surround filter. Connected regions yielding high responses to this filter are then checked for size. Regions which are not within allowable nose dimensions are eliminated. The maximum allowable height is about 75mm and the maximum allowable width is about 35mm. Final ridge candidates are constrained to overlap one of the regions satisfying this protrusion test. Future implementations will also test for proximity to the symmetry line. Nose identification is currently a critical link in the system, because the eye search is based on its results.

Consistent detection and measurement is a desirable property in selecting features to describe aspects of the face. One of the more consistent curvature features is evident in figure 2. There are sets of symmetric concavities (black blobs) which occur at the inside corners of the eyes or nose bridge, the outside corner of the eyes, and the base of the nose. They also often occur at the nostrils, the outside corners of the mouth, and at dimples. Of these points, the corners of the mouth and the dimples often have a lower degree of symmetry than the rest. Symmetric convex regions (white blobs) are also evident; the pair that is most consistently present defines the location of the eyeballs (or eye lids). Our system finds the eye corners and centers using these features. We begin with an image of

Feature	Geometric constraints			Symmetry constraints		Proximity Target	Verification Conditions
	area	width	height	dist. limits $d_s$	asymm. $(\Delta d_s, \Delta p_s)$		
eyeballs (eb)	0 - 350	<u>0 - 40</u>	<u>0 - 17</u>	35 - 70	12, 15	$ p_s - p_s(\text{nose center})  < 40$	<ul style="list-style-type: none"> <li>• <math>p_s</math> above <math>p_s(\text{nose base})</math></li> </ul>
inside corner eye cavity (iec)	<u>150</u> - 900	0 - 300	0 - 300	10 - 25	12, 15	$ p_s - p_s(\text{eb})  < 14$ or $ p_s - p_s(\text{nose center})  < 40$	<ul style="list-style-type: none"> <li>• <math>p_s</math> above <math>p_s(\text{nose base})</math></li> <li>• <math>d_s(\text{inside corner}) &lt; d_s(\text{eyeball})</math></li> </ul>
outside corner eye cavity (oec)	0 - 900	0 - 300	0 - 300	40 - 70	12, 15	$ p_s - p_s(\text{eb})  < 14$ or $ p_s - p_s(\text{iec})  < 17$ or $ p_s - p_s(\text{nose center})  < 30$	<ul style="list-style-type: none"> <li>• <math>p_s</math> above <math>p_s(\text{nose base})</math></li> <li>• <math>d_s(\text{outside corner}) &gt; d_s(\text{eyeball})</math></li> <li>• <math>d_s(\text{inside corner}) &gt; d_s(\text{inside corner})</math></li> </ul>

*Figure 5 : Summary of constraints used in selection of eye descriptors. All values are given in mm (or mm<sup>2</sup> for area). Underlined geometric constraints represent actual limits, whereas other values are currently large enough to represent don't care conditions.*

connected concave or convex regions; the location of each connected region is defined by its center of mass. Geometric constraints based on allowable distances and sizes for eyes are used first to limit the search, then symmetry constraints are used to locate candidate pairs. If there are multiple pairs which satisfy both symmetry and geometric constraints, the selection is made based on closest distance between the pair's projection onto the symmetry line and a given target. The target is on the symmetry line, and is usually given in terms of the position of other known features (in the first trial only the nose is defined). The selected location is also tested for validity against other features' relative positions (eg. outside corners of the eyes must be farther from the symmetry line than the inside corners of the eyes). The verification can also continue as features become defined. Constraints are summarized in figure 5. The symmetry constraints shown are the allowable range of distances from the point to the symmetry line ( $|d_s|$ ), and the allowable asymmetry between matched pairs where  $p_s$  is the projection of a point onto the symmetry line. An example of the search (for the same face shown in figure 2a) is summarized graphically in figure 6; the entire connected component image is given in light grey, blobs satisfying geometric constraints are marked in dark grey, blobs satisfying symmetry matching constraints are connected, and the selected pair is marked in black.



*Figure 6 : (a) selection of the convex region marking the eyelid/eyeball, (b) selection of the concave region at the inside corner of the eye, and (c) selection of concave region at the outside corner of the eye. Example face is the same as shown in figure 3a.*

Some features are best defined in terms of regions. Hoffman and Richards defined a rule for partitioning a surface into parts in a manner which was supposed to be consistent with the way we naturally mentally segment or organize a shape [Hoffman and Richards 1984]. They suggest to “divide a surface into parts at loci of negative minima of each principal curvature along its associated family of lines of curvature.” The applicability of this rule to complex objects has not previously been tested, particularly on real data. One might hope that features which

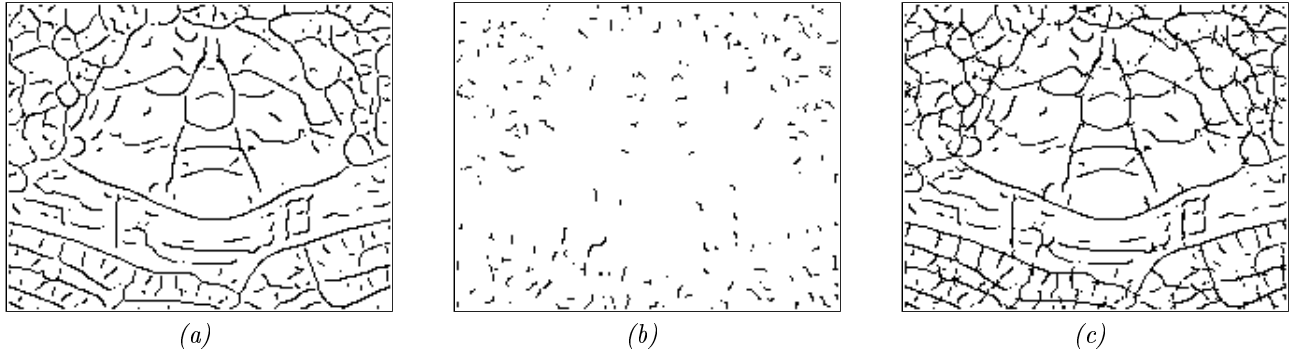


Figure 7: Points on part boundaries in (a)  $k_{min}$ , (b)  $k_{max}$ . (c) complete part boundary image.

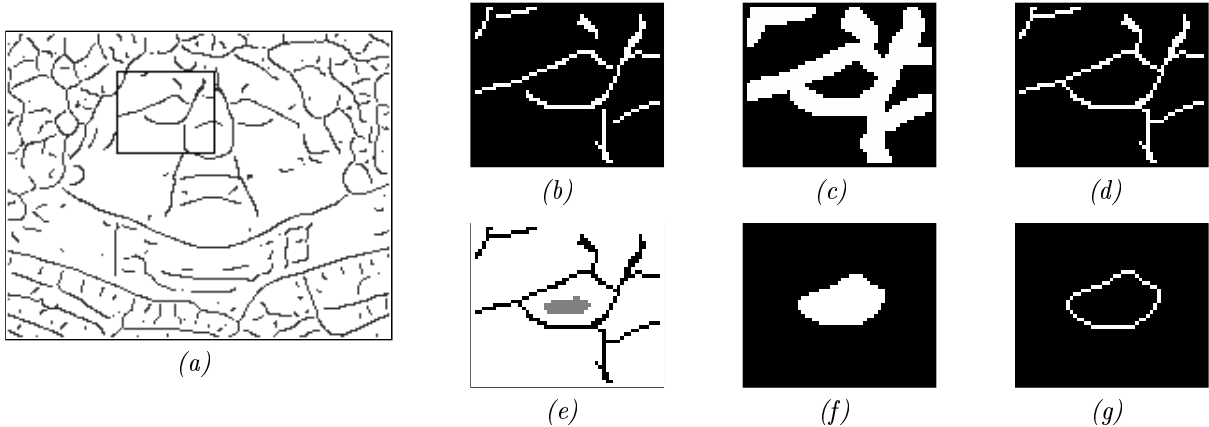
most of us naturally segment as “parts” of the human face such as the nose, mouth, or eyes might be partitioned effectively by this rule. It is a simple matter to evaluate this part boundary criterion locally; a point is said to fall on a part boundary if it is at a negative local minimum in either principle curvature along its respective direction vector. As in ridge and valley finding this is evaluated using bilinear interpolation on the  $\theta, y$  lattice as determined by the principal direction vector. In fact the boundary points along the lines of minimum curvature are exactly the same as the points we have called valley lines in the previous section; the threshold in this case is specified as zero, where we usually choose a smaller value. Some results of this test are shown in figure 7.

Although close in some cases, it is important to note that this algorithm does not generally yield completely connected boundaries. It is striking that even the nose, which is closest to the Hoffman and Richard’s ideal of a convex protrusion added to a lower curvature surface by a smooth join does not appear as a complete part. A major problem is that, unlike the zero crossings of a function, their definition of part boundaries creates curves which can end abruptly with no continuation. This occurs when a local minimum and a local maximum on one line of curvature coalesce and disappear in nearby lines of curvature. Combined with *a priori* knowledge of the face, however, this information can become very useful. If we know what we are looking for we can use this information to help fill in the gaps and identify significant boundary points using some common morphological operators. We give as an example the case of locating bounding areas for the eye socket and nose.

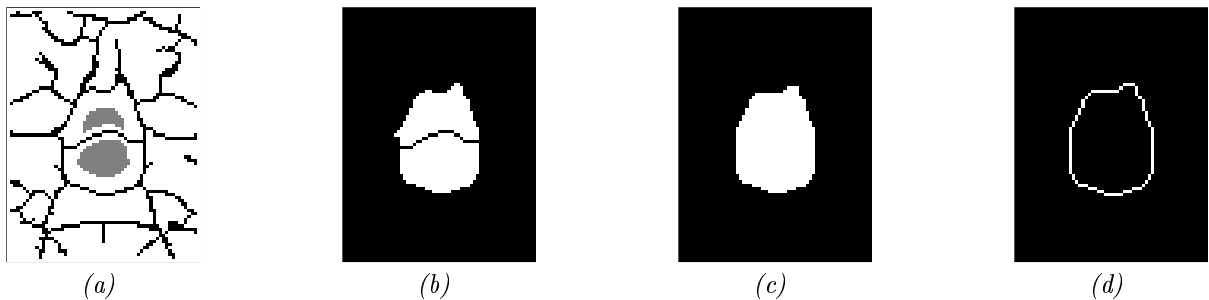
We first use the location of the corner cavities of the eyes (or approximations based on the eyeball location if the cavities weren’t located successfully) to establish a bounding area for the search. An example is shown by the rectangle superimposed on figure 8a. The first of the small images shows the boundary points found in the subimage (8b). We dilate the image, bridging the important gaps (8c). Then we compute the skeleton of the enlarged region, using the constraint that the points in the original boundary image must be included in the skeleton (8d) [Vincent 1991]. We invert the skeleton and use the convex region identified as the eyeball or lid as a seed for region growing (8e); this selects the connected region belonging to the eyesocket(8f). We can use either the region or its connected and isolated boundary to describe this area.

Figure 9 shows a similar procedure to identify the nose region. The seeds for region growing are determined by convex regions on the nose ridge. It is common to find more than one such region because there is often at least one “dip” in the profile of the nose ridge. Figure 9b shows regions selected by both seeds; figure 9c shows these regions after a morphological closing which joins and smoothes the regions. A graphical summary of these eye and nose feature points and regions as located for several sample faces is shown in figure 10.

As discussed earlier, it is important to know what part of the surface actually corresponds to the face. Here we can use the fact that the skin is relatively smooth in comparison with hair or clothing boundaries. We create a binary image using the criteria  $k_{min}^2 + k_{max}^2 < thresh$ . The resulting mask will include the smooth portions of the face on the forehead, cheeks, and neck, as well as small separate regions in hair or clothing. It will not, however, include the eyes, nose, or mouth, or other places on the face where interesting curvature events occur. But since we have already found the eye and nose regions, we can add them to the mask (we first dilate them for better coverage). We do not attempt to cover holes in the mask due to the mouth and chin, because these areas have a high degree of variation caused by expression and position and we would like to minimize these differences. In the future the



*Figure 8: Sequence of images illustrating the calculation of the eye region*



*Figure 9: Sequence of images illustrating the calculation of the nose region*

elimination of these areas in the mask will be considered in a more formal way. The skin region is separated from the rest of the image (which may include small spurious regions) by selecting the connected component with the largest area. This mask defining the face region is used to decide over what points valid comparisons can be made between faces.

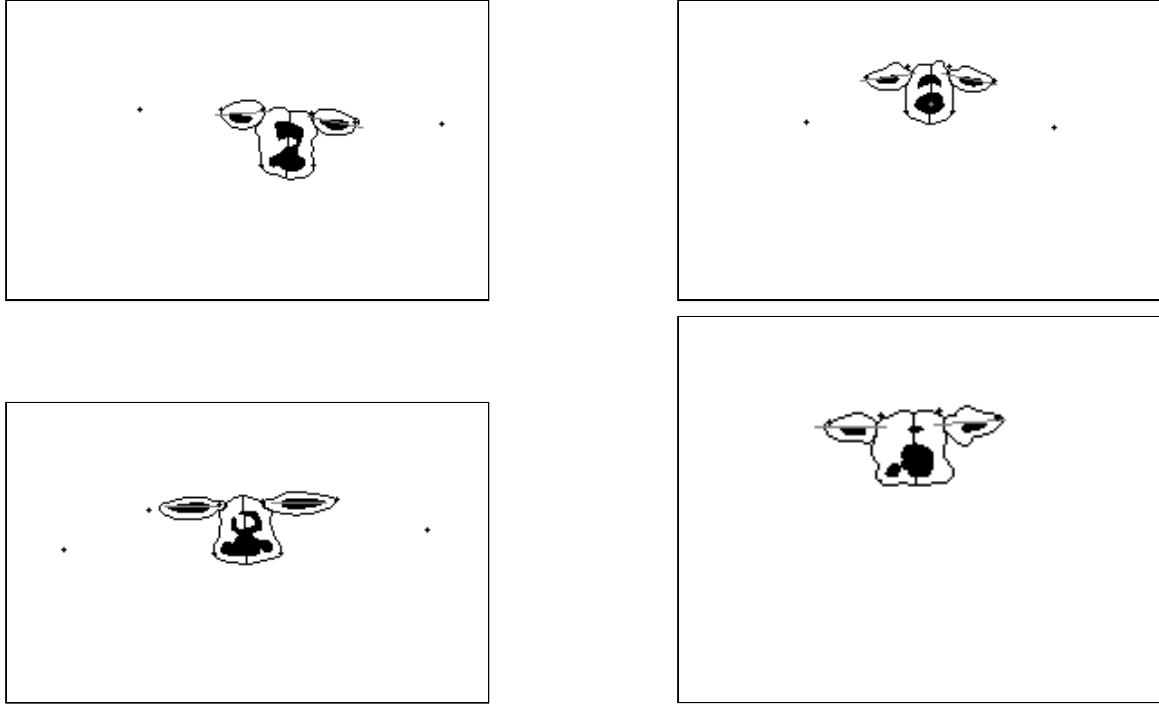
### 3.2 Implementation of depth template comparison

Our aim is to define a transformation of the face to a normalized position which is consistent for an individual independent of view or expression. We use the location of the bridge of the nose, the base of the nose, and the outside corners of the eyes. The line connecting the two nose points is aligned with the vertical or y axis (determines the rotation about the x and z axis), and the line determined by the two outside corners of the eyes is aligned roughly parallel to the x axis (determines rotation about y axis). The nose bridge is positioned at a fixed translation from the origin.

Once the surface is transformed to the standard position it is reinterpolated onto the regular cylindrical grid. This is accomplished with a cylindrical ray tracing algorithm. We then compute the volume of the space between the two normalized surfaces as a measure of their similarity; small values indicate a high degree of similarity. For two faces,  $a$  and  $b$ , this volume can be approximated by

$$\text{diff}(a, b) = \sum_{\text{valid } (\theta, y)} |R_a^2(\theta, y) - R_b^2(\theta, y)|, \quad (5)$$

where  $R_a(\theta, y)$  is radius at  $(\theta, y)$  for surface  $a$ , and similarly for surface  $b$ . The mask specifying which points are to be considered part of the face is also mapped to the normalized position, and comparisons are only made in the intersection of the two valid regions. This criteria can be used to judge similarity between a test face and each face in a data base. The comparison yielding the smallest difference is identified as the best match to the test face.



*Figure 10 : Sample results of feature detection.*

### 3.3 Experimental Results

There are two stages involved in performing recognition experiments. First each face in the data base must undergo feature classification. Then comparisons can be made between all faces for which feature detection was sufficient to allow the construction of the face area mask and normalization of the face to the standard position.

A training set which consisted of 26 individual faces was used to develop and tune the feature detection portion of the system. The faces included 8 women and 18 men; of this group no one wore eyeglasses, moustaches or beards. The range of ages of the subjects was roughly from early 20's to late 50's. Little information was available about ethnic origin. Feature detection was rated *sufficient* for normalization if the verification conditions were satisfied for the outlines of eye and nose regions, the nose bridge and base, and the inside corner eye cavities. Alternatively if the eye outlines were not successfully identified, they could be approximated if the outside eye corners were identified. So this condition also was considered sufficient. This sufficiency decision is automated and does not guarantee that the feature recognition was accurate. Sufficient feature detection was achieved for 100% of the training set faces.

A *separate* test set was used in recognition experiments. The test set consisted of 8 faces with 3 views each for a total of 24 faces. For each face there are two versions without expression (that is with a neutral expression), and one with an expression. In this group there were 4 women and 4 men. Sufficient feature detection was achieved for 100% of these faces.

The 24 faces in the test set were then subjected to depth template matching. One view of each face was compared against the remaining 23 faces. The faces were ranked in order of increasing average difference. In 8 out of 8 cases the first ranked image was another view of the same subject. For identification purposes we could consider this a 100% recognition rate. However, there are two faces in the data base matching the identity of each target subject, and in only 5 out of the 8 cases was the second ranked face another version of the target.

To be more specific about these results let us formalize our recognition hypothesis. For this process to be successful we would like the difference between two instances of the same face to be smaller than the difference between two faces of different subjects. If we have  $m$  subjects, and  $n$  instances of each subject, for a given individual

we can divide the total number of faces into two sets:

$$A = (\alpha_1, \dots, \alpha_n), \tag{6}$$

$$B = (\beta_1, \dots, \beta_{(m-1)n}). \tag{7}$$

Set  $A$  contains all instances of the individual, and set  $B$  contains all instances of other subjects. Then we hypothesize that for  $i \neq j$

$$\text{diff}(\alpha_i, \alpha_j) < \text{diff}(\alpha_i, \beta_k), \tag{8}$$

where  $i, j < n$  and  $k < (m-1)n$ . The total number of comparisons of this type is  $m((n-1)(m-1)n)$ , or for our case ( $m = 8$ , and  $n = 3$ ), 336. These tests provides a better evaluation of the entire ranking. It takes into account how far off the ranking was if a recognition error occurs. For instance, if we are attempting to match two instances of the same face and 10 faces of different subjects are ranked higher than the other instance of the target face, this will count as 10 incorrect comparisons. We found that in 11 cases out of 336 the hypothesis was not satisfied. So 97% of the comparisons were correct.

## 4 Conclusions

This paper is intended to give an outline of the use of curvature information in the process of face recognition. We have discussed many advantages this approach has over intensity based methods first in its ability to describe several types of structures in the face such as cheeks, forehead, and chin, and second in that fact that surface curvature is invariant with respect to viewing angle. The computation of a plethora of useful surface primitives is presented and their value is discussed in the context of face recognition. We have also demonstrated examples of face specific feature detection, and their use in a simple brute force depth comparison recognition strategy. The results are very encouraging with excellent recognition rates demonstrated in this implementation. We can further improve the system by formally eliminating the consideration of the neck area, which can have a large position variation relative to the face, and the mouth area which shows the largest change due to facial expression.

There is a great deal of information contained in the curvature maps which we haven't yet taken advantage of explicitly. Ridges in  $k_{max}$  would help us to identify several other obvious structures of the face such as chin, and brow line, and to verify the position of the nose ridge. We consider the shape of cheek and forehead implicitly on a point by point basis; however, we could use the range and structure of the curvature values within certain regions of the face to develop specific descriptors. These features would correspond to quantitative versions of Harmon's descriptors for shape of cheeks ("sunken" or "hollow") and forehead ("receding" or "bulging"). With these additional feature descriptors a comparison of faces directly in feature space becomes possible as an alternative recognition strategy.

## Acknowledgments

I'd like to thank David Mumford, Alan Yuille, Taka Shiota, Peter Hallinan, Mark Nitzberg, and Luc Vincent for their valuable discussions and advice during the course of this work. The range data presented in this paper was generated by a Cyberware Laboratory 4020/PS 3-D Digitizer. This work was supported in part by the U.S. Army Research Office under grant DAAL03-86-K-0171.

## References

1. P. F. Besl and R. C. Jain, "Invariant Surface Characteristics for 3D Object Recognition in Range Images". *Computer Vision, Graphics, and Image Processing* 33,33-80 (1986).
2. P.J. Besl and R.C. Jain, "Three-dimensional Object Recognition," *ACM Comput. Surveys*, Vol 17, No. 1, 1985, pp. 75-145.
3. Davies, Ellis, Shepherd. "Face recognition accuracy as a function of mode of representation", *Journal of Applied Psychology*, vol. 73, 105-116, 1978.
4. T.J. Fan, G. Medioni and R. Nevatia, "Description of Surfaces from Range Data Using Curvature Properties." *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 86-91, 1986.

5. M.A.Fischler, and R.A.Elschlager, "The Representation and Matching of Pictorial Structures", *IEEE Transactions on Computers*, Vol. c-22, No. 1, January 1973. (Lockheed Palo Alto Research Lab, Lockheed Missiles and Space Company LMSC-D243781 Sept. 1971).
6. G. G. Gordon, "Smoothing Range Data for Curvature Estimation," *Active Perception and Robot Vision, Proceedings of the NATO Advanced Study Institute, Maratea, Italy, July 1989*, A.K.Sood, H.Wechsler eds. Springer Verlag, to appear 1991.
7. G. G. Gordon, "Feasibility Study: Face Recognition from Surface Curvature Data," Harvard Robotics Laboratory Technical Report, November 1990.
8. P.W. Hallinan, "Recognizing Human Eyes," Harvard Robotics Laboratory Technical Report, November 1990.
9. Harmon, L.D. "The Recognition of Faces", *Scientific American*, vol 229, 1973.
10. D.D. Hoffman and Whitman Richards, *Parts of Recognition*. MIT Artificial Intelligence Laboratory Memo No. 732, December 1983
11. T.Kanade, "Picture Processing System by Computer Complex and Recognition of Human Faces," Doctoral Thesis, Department of Information Science, Kyoto University, 1973.
12. J. Lee, E. Miliou, "Matching Range Images of Human Faces," International Conf. Computer Vision, Dec. 1990.
13. N. Masui, S. Akamatsu, Y. Suenaga, "A Preliminary study for Recognition of Human Faces by 3-D Measurement", ICS 90-51, AIPS 90-43 June 1990 (In Japanese).
14. R. S. Millman and G. D. Parker, *Elements of Differential Geometry*. New York: Prentice-Hall, 1977.
15. Rhodes, Brennan, Carey. "Identification and Ratings of Caricatures: Implications for Mental Representations of Faces", *Cognitive Psychology* 1987 vol. 19, 473-497.
16. Sander, and Zucker, "Singularities of Principal Direction Fields from 3-D Images," in press IEEE, PAMI.
17. M. Turk, and S. Pentland, "Eigenfaces for Recognition." *Journal of Cognitive Neuroscience*, March 1991.
18. B.C. Vemuri, Mitiche, and J.K. Aggarwal, "Curvature-based representation of objects from range data," *Image and Vision Computing*, Vol. 4, No. 2, May 1986.
19. L. Vincent, "Efficient computation of various types of skeletons," *Proc. SPIE Medical Imaging V*, San Jose, CA, February 1991.
20. A. L. Yuille, D. S. Cohen, and P. W. Hallinan, "Feature extraction from faces using deformable templates," CVPR 1989.