

Face Recognition Based on Depth and Curvature Features

Gaile G. Gordon *

TASC

55 Walkers Brook Drive

Reading, MA 01867

Abstract

This paper explores face recognition from a representation based on features extracted from range images. Depth and curvature features have several advantages over more traditional intensity based features. Specifically, curvature descriptors 1) have the potential for higher accuracy in describing surface based events, 2) are better suited to describe properties of the face in areas such as the cheeks, forehead, and chin, and 3) are viewpoint invariant. Faces are represented in terms of a vector of feature descriptors. Comparison between two faces is made based on their relationship in the feature space. We provide detailed analysis of the accuracy and discrimination of the particular features extracted, and of the effectiveness of the recognition system for our test database of 24 faces. Results are very promising. In many cases it is shown that feature accuracy is limited more by surface resolution than by the extraction process. Recognition rates in our experiments are in the range of 80% to 100%.

1 Introduction and Background

The task of automated face recognition involves first the development of a quantitative representation of the face, and second the comparison and matching of faces in terms of this representation. A general purpose face recognition system must be able to distinguish between a large number of people under typical kinds of image variation, for example: lighting conditions, viewing position, facial expressions, or changes in superficial detail such as hair style. We argue that a recognition paradigm based on range input shows more potential in the development of a general purpose system than approaches which use features extracted from intensity images [1, 2]. Previous work [3] presented a face recognition system based on range data template matching. This paper presents a face recognition system which uses *features* extracted from range and curvature data to represent the face (see [4] for a description of curvature calculation from range data). Features are extracted in two stages. First, high level features are identified in terms of points, lines, and regions on the surface. In the current implementation, these high level features describe aspects of

the eyes, nose, and head. Second, low level scalar features are extracted in terms of distance or curvature measurements. The set of measurements for each face describes a single point in feature space. Comparison between two faces is made based on their relationship in feature space.

An approach based on range and curvature data has several advantages over intensity image approaches. In the context of a recognition system, the requirements for accuracy in a specific feature are dependent on the number and accuracy of the other features used. In general, the larger the feature base, the less strict the accuracy requirements. In past approaches, the features used in describing faces have been limited to the eyes, nose, mouth, and face boundary. The additional information contained in low contrast areas of the face (the jaw boundary, cheeks, and forehead) has been ignored because these areas are difficult to describe reliably from intensity images. If we could extract information about these regions it would greatly increase the size of the available feature set. Curvature measures, which can be calculated from accurate surface data, are an obvious choice for this task. Also, as a local surface property of the object, curvature has the valuable characteristic of being *viewpoint invariant*. Thus, depth based processing has the potential for better accuracy, a richer feature base, and provides the potential to use curvature measures which are viewpoint invariant.

This paper presents a summary of the face descriptors used, the recognition strategy, and experimental results. For more detail please consult [3, 6, 4].

2 Face Descriptors

The first stage in the description of a face is to extract high level features which mark salient events on the face surface in terms of points, lines, and regions. This process provides a segmentation of the face data. The current implementation focuses on eyes, nose, and head and includes the following features:

- nose bridge (nasion),
- nose base (base of septum),
- nose ridge,
- eye corner cavities: inner and outer,
- convex center of the eye (eyeball/lid region),
- eye socket boundary,
- boundary surrounding nose,

*This work was performed primarily at The Harvard Robotics Laboratory and was supported in part by the U.S. Army Research Office under grant DAAL03-86-K-0171.

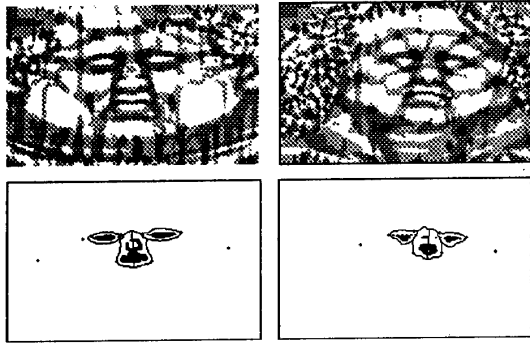


Figure 1: Two faces segmented by (top) sign of Gaussian and mean curvature, and (bot) high level descriptors as discussed: eye-socket and nose region outlines with their markers, other feature points marked by crosses. (Shown in cylindrical coord.s (θ, y)).

- opposing positions on the the cheeks for measurement of head width.

Each of these face descriptors is defined in terms of a high level set of relationships of depth and curvature values. As such, the feature extraction process is structured as a constrained search on the surface. As each constraint is enforced, the number of potential locations for the feature is reduced. The constraints are designed to reduce the search to a single definition of the feature. The search process considers constraints on:

- sign of Gaussian and mean curvature
- absolute extent of a region on the surface
- distance from the symmetry plane
- symmetrical pairing of features
- proximity to a target on the surface
- protrusion from the surrounding surface
- local configuration of curvature extrema

Examples of these features as extracted for four different faces are shown in figure 1.

A set of scalar features is calculated from these high level point and region features. The most basic scalar features correspond to measurements of distance and curvature. From the above descriptors we have chosen some simple features:

- left and right eye width
- eye separation
- total span of eyes
- nose height, width, and depth
- head width.
- maximum Gaussian curvature on the nose ridge
- average minimum curvature on the nose ridge
- Gaussian curvature at the nose bridge and base

In case of distance measures we use an absolute scale (mm) not a relative scale (pixels). Thus, we are not restricted to using ratios of distances as in the case of measurements calculated from intensity images.

3 Comparing Faces

The vector formed by the set of descriptor values for a given face places the face in the space of all possible faces. To

use this representation as an effective basis for recognition we require that the all the points in feature space which correspond to the same person will cluster with regard to some similarity measure.

There is a prolific literature on statistical methods for clustering analysis (a good introduction is found in [7]) which includes a thorough examination of similarity measures and cluster segmentation. For the purposes of this investigation, however, we will use the Euclidean distance measure in scaled feature space. We consider this analysis as a feasibility study which will provide a good baseline for future work.

4 Experimental Results

A training set which consisted of 26 individual faces was used to develop and tune the feature detection portion of the system. A *separate* test set was used in recognition experiments. This 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 a different facial expression from the neutral versions.

Our results are divided into two sections; first is the evaluation of the individual feature descriptors, and second is the evaluation of the recognition system itself.

In evaluating the usefulness of features in automatic face recognition we consider two criteria. First they must be robustly detectable; their measurement must be consistent for the same face over reasonable variation in view position, expression, age, weight, etc. Second their values must vary distinctly over the range of different individuals. We present a statistical evaluation of our features with respect to both these factors, repeatability and distinctiveness. Since our data base is relatively small, statistical evaluation can not always be a good characterization of the results. A more detailed evaluation, and the raw feature data itself is included in [4].

First we address repeatability. For each feature we calculated the range of that feature value for each of the 8 subjects. From among these 8 ranges we examined the minimum variation, the maximum variation, and the median variation. The median variation statistic is the most representative of the repeatability we can expect from the feature extraction; the maximum within subject variation is high in a few case, but this is due to specific problems in one or two individual measurements [4]. For the distance measures the median variation within subject ranged from about 1.4mm to 5.5mm for different features, with the majority falling between 2 and 3mm. We show that the feature extraction seems limited primarily by the resolution of the range data.

How valuable a feature is in discrimination among subjects is a function both of repeatability within subject and variation between subjects. We can evaluate this property numerically for our particular data base by using a multiple class version of Fisher's linear discriminant criterion [7]. This criterion evaluates the discriminating power of a given feature by considering the ratio of between class variance

| Feature set considered | Recognition Rates | Correct Ranking Rates |
|------------------------|-------------------|-----------------------|
| I | 70.8 | 96.1 |
| | 75.0 | |
| II | 91.7 | 98.1 |
| | 83.3 | |
| IV | 95.8 | 98.2 |
| | 79.2 | |
| III | 100 | 99.4 |
| | 79.2 | |

Table 1: Recognition rates and correct ranking rates.

to within class variance. Higher values of this criterion indicate better discriminating power.

Using this criterion [6] we find there is quite a range in the discrimination power of our 12 features (the best was rated 24.9, the worst 1.3). We can use this ranking to determine which features to consider further (or weight more highly) in the comparison process. We must note that the usefulness of a feature in discrimination is by nature a function of the particular data base considered.

Next we evaluate the comparison method itself. In our experiments, for each target face, the other faces in the data base are sorted by decreasing similarity. We evaluate this ordering or ranking of the data base for each target in two different ways. First, we consider a traditional "recognition rate" evaluation, and second we evaluate the correctness of the entire ranking for each target. The comparison process is obviously sensitive to which features we choose initially to represent the face since this defines the feature space. As we have shown above there is wide variation in the discriminating power of our features. We used four different combinations of features. The basic set, denoted (I), includes the best 4 features head width, nose height, depth, and width. The other three sets include increasing numbers of features added in order of discriminating power. We will define each of our evaluation methods, report the results for each, and finally discuss the implications of these results.

Correct recognition of a target requires that the most similar face in the data base (the face with smallest distance from the target) has the same identity as the target. A more stringent recognition rate evaluation requires that *all* faces in the data base with the same identity as the target are found to be more similar to the target than other faces in the data base. For each of the targets there are two faces with the same identity remaining in the data base. Table 1 shows for each feature set we considered, the percentage of targets for which the best match was correct (top), and the percentage of targets for which the second best match was also correct (bottom).

We can formalize this recognition criterion further to evaluate the entire rank for each target. For successful recognition we hypothesize that the difference between two instances of the same face should be smaller than the difference between two faces of different subjects. We evaluated this hypothesis formally [6]. Table 1 shows our results in terms of percentage of correct rank orderings out of 1008.

5 Conclusions

These results show certain trends. Performance improved with additional features (I, II, III) until the discrimination power of the new feature (IV) was too low to contribute positively to recognition. At this point the performance will actually decrease. This is because for features with a low discrimination rating, the feature values are all very similar, *except for the cases with large feature detection error, or variation due to expression*. Thus, the only discrimination of the new feature is on the basis other than identity. As long as the features are above some threshold in the discrimination rating we can predict continued improvement for larger feature sets.

The raw recognition results for the closest match condition showed no errors out of 24 cases using the optimal feature set. This indicates that clustering with respect to the Euclidean distance measure is occurring. However, there were more errors in identifying the second instance of the target face in the data base (19 or 20 cases correct out of 24). The evaluation of the relative ranking of the data base members with respect to each target gives us a better picture of performance. With the optimal feature set there were only 6 ordering errors out of 1008 comparisons in the ranking. A low number of errors indicates the correct matches are closer to the top of the ranking - which means the clustering is better. However, the distances between unrelated faces is similar in some cases to the distances between related faces, so there is not always a distinct separation between the clusters.

The performance of this system shows great promise. Even with a small feature set and very basic statistical methods for classification, raw recognition results were at 100% with the best feature set, and there were only 6 errors in the ranking of the data base over all target matches. These experiments clearly demonstrated the usefulness of the features developed in distinguishing among faces. Further work on improving statistical comparison process and consideration of additional features is in progress.

References

- [1] T. Kanade, *Picture Processing System by Computer Complex and Recognition of Human Faces*. PhD thesis, Kyoto University, 1973. Department of Information Science.
- [2] A. L. Yuille, D. S. Cohen, and P. W. Hallinan, "Feature extraction from faces using deformable templates," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (San Diego, CA), pp. 104-109, June 1989.
- [3] G. G. Gordon, "Face recognition from depth maps and surface curvature," in *Proceedings of SPIE Conference on Geometric Methods in Computer Vision*, (San Diego, CA), July 1991.
- [4] G. G. Gordon, *Face Recognition from Depth and Curvature*. PhD thesis, Harvard University, Division of Applied Sciences, 1991.
- [5] M. A. Turk and A. P. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.
- [6] G. G. Gordon, "Face recognition using depth and curvature features," Technical Report 91-23, Harvard University Robotics Lab, Cambridge, MA, December 1991.
- [7] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley and Sons, 1973.