Multiple Nose Region Matching for 3D Face Recognition under Varying Facial Expression

Kyong I. Chang, Kevin W. Bowyer, *Fellow*, *IEEE*, and Patrick J. Flynn, *Senior Member*, *IEEE*

Abstract—An algorithm is proposed for 3D face recognition in the presence of varied facial expressions. It is based on combining the match scores from matching multiple overlapping regions around the nose. Experimental results are presented using the largest database employed to date in 3D face recognition studies, over 4,000 scans of 449 subjects. Results show substantial improvement over matching the shape of a single larger frontal face region. This is the first approach to use multiple overlapping regions around the nose to handle the problem of expression variation.

Index Terms—Biometrics, face recognition, three-dimensional face, facial expression.

✦

1 INTRODUCTION

FACE recognition using 3D shape is believed to offer advantages over the use of intensity images [1], [2], [3]. Research on face recognition using 3D shape has recently begun to look at the problem of handling the variations in shape caused by facial expressions [4], [5], [6], [7], [8]. Various approaches might be employed for this purpose. One is to concentrate on regions of the face whose shape changes the least with facial expression [9], [10]. For example, one might ignore the lips and mouth, since their shape varies greatly with expression. Of course, there is no large subset of the face that is perfectly rigid across a broad range of expressions. Another approach is to enroll a person into the gallery using a set of different expressions. However, the probe shape may still be an expression different than those sampled. A third approach is to have a model of 3D facial expression that can be applied to any face shape. However, there likely is no general model to predict, for example, how each person's neutral expression is transformed into their smile. A smile is different for different persons and for the same person at different times. A fourth approach is to try to compute an expression-invariant representation of the 3D face shape [11], [12].

Given that there is no fully "correct" approach to handling varying facial expression, one question is which approach(es) can be most effectively used to achieve desired levels of performance. In this work, we explore an approach that matches multiple, overlapping surface patches around the nose area and then combines the results from these matches to achieve greater accuracy. Thus, this work seeks to explore what can be achieved by using a subset of the face surface that is approximately rigid across expression variation.

2 BASELINE PCA AND ICP PERFORMANCE

We first establish "baseline" performance levels for 3D face recognition on the data set used in this work. Images are obtained using a Minolta 900/910 sensor that produces registered 640×480 range and color images. The sensor takes several seconds to acquire the data and subject motion can result in

---

- *K.I. Chang is with Philips Medical Systems, 22100 Bothell Everett Hwy, Bothell, WA 98021. E-mail: Jin.Chang@philips.com.*
- *K.W. Bowyer and P.J. Flynn are with the Department of Computer Science and Engineering, University of Notre Dame, 384 Fitzpatrick Hall, Notre Dame, IN 46556. E-mail: {kwb, Flynn}@cse.nd.edu.*

artifacts [7]. Images with noticeable artifacts result in recognition errors. See Fig. 1 for examples of the various facial expressions. For the baseline algorithms, we use a PCA approach similar to previous work [1], [13], [14] and an iterative closest point (ICP) approach similar to previous work [2], [10], [15]. More sophisticated approaches have appeared in the literature [3], [4], [5], [6], [7], [8]. These "baseline" approaches are simply meant to represent common known approaches. See Fig. 2 for examples of the frontal face regions used for these baseline algorithms. A total of 546 subjects participated in one or more data acquisitions, yielding a total of 4,485 3D scans as summarized in Table 1. Acquisition sessions took place at intervals over approximately a year and, so, a subject may have changes in hair style, weight, and other factors across their set of images. Among the 546 subjects, 449 participated in both a gallery acquisition and one or more probe acquisitions. The earliest scan with a neutral expression is used for the gallery and all later scans are used as probes. The neutral-expression probe images are divided into nine sets, based on increasing time lapse between acquisition sessions. There are 2,349 neutral-expression probes, one or more for each of the 449 neutral-expression gallery images. The nonneutral-expression probe images fall into eight sets, based on increasing time lapse. There are 1,590 nonneutral probes, one or more for each of 355 subjects with neutral-expression gallery images.

Results for the PCA baseline are created using manually-identified landmark points to register the 3D data to create the depth image. The training set for the PCA algorithm contains the 449 gallery images plus the 97 images of subjects for whom only one good scan was available. Results for the ICP baseline use the manually-identified landmark points to obtain the initial rotation and translation estimate for the ICP matching. In this sense, the baseline represents an idealized level of performance for these approaches. There is a significant performance decrease when expression varies between gallery and probe, from an average 91 percent to 61.5 percent for the ICP baseline, and from 77.7 percent to 61.3 percent for the PCA baseline. The higher performance obtained by the ICP baseline is likely due to the fact that ICP handles pose variation between gallery and probe better than PCA. These results agree with previously reported observations: one, that ICP approaches outperform PCA approaches for 3D face recognition [10], [16], and, two, that expression variation degrades recognition performance [4], [5], [6], [7], [8], [17].

3 MULTIPLE NOSE REGION MATCHING

Beginning with an approximately frontal scan, the eye pits, nose tip, and bridge of the nose are automatically located. These landmarks are used to define, for a gallery face shape, one larger surface region around the nose; and for a probe face shape, multiple smaller, overlapping surface regions around the nose. For recognition, the multiple probe shape regions are individually matched to the gallery and their results combined.

3.1 Preprocessing and Facial Region Extraction

Preprocessing steps isolate the face region in the scan. Considering the range image as a binary image in which each pixel has or doesn't have a valid measurement, isolated small regions are removed using a morphological opening operator (radius of 10 pixels). Then, connected component labeling is performed, and the largest region is kept; see Fig. 3. Outliers, which can occur due to range sensor artifacts, are eliminated by examining the variance in Z values. A 3D point is labeled as an outlier when the angle between the optical axis and the point's local surface normal is greater than a threshold value (80 degrees). Next, the 2D color pixels corresponding to these 3D points are transformed into YCbCr color space and used for skin detection. The 3D data points in the detected skin region are subsampled keeping the points in every fourth row and column. (On average, there are more than 11,000 points in the face region, about 3,000 points in the gallery surface, and 500 to 1,000 points in the probe surfaces.) This reduces computation in later steps and
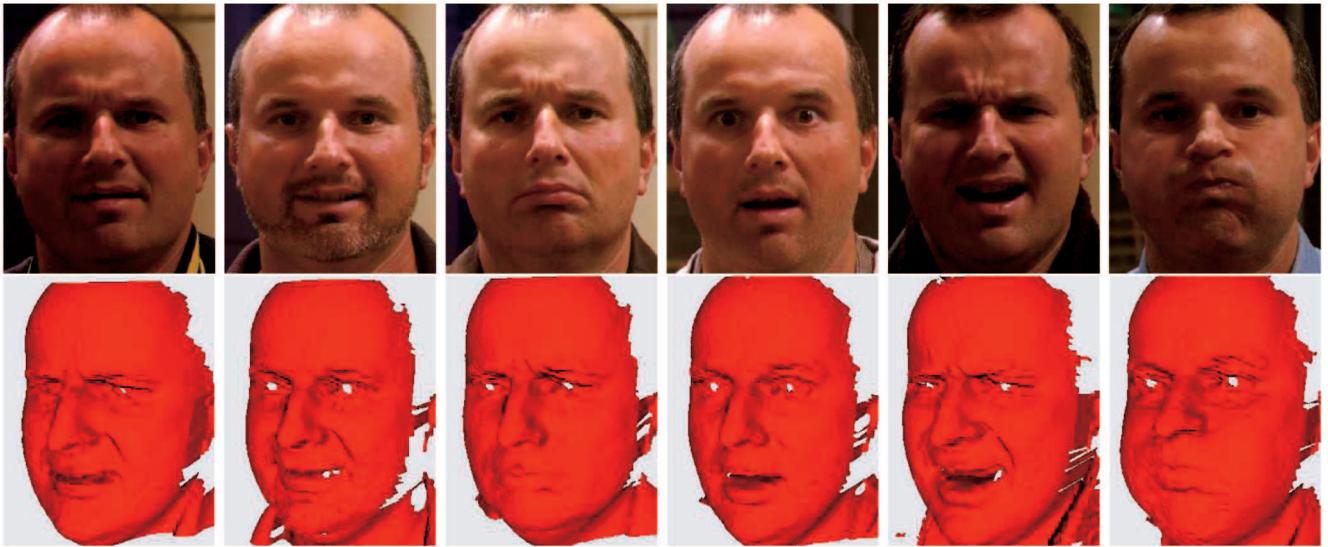
Fig. 1. Example images in 2D and 3D with different facial expressions.

initial smaller experiments indicated that it does not significantly affect recognition.

3.2 Curvature-Based Segmentation and Landmark Detection

We compute surface curvature at each point, create a region segmentation based on curvature type, and then detect landmarks on the face. A local coordinate system for a small region around each point is established prior to the curvature computation, formed by the tangent plane (X-Y plane) and a surface normal (Z axis) at the point. Using a PCA analysis of the points in the local region, the X and Y axes are the eigenvectors of two largest eigenvalues and Z axis is the smallest eigenvector, assumed to be the surface normal. Once the local coordinate system is established, a quadratic surface is fit to the local region. After the coefficients for the fit are found, partial derivatives are computed to estimate mean curvature, H, and Gaussian curvature, K. The curvature type is labeled based on H and K and points with the same curvature type are grouped to form regions.

Fig. 4 illustrates the steps to detect the nose tip (peak region), eye cavities (pit region), and nose bridge (saddle region). A nose tip is expected to be a peak ($\mathbf{K} > T_K$ and $\mathbf{H} < T_H$), a pair of eye cavities to be a pair of pit regions ($\mathbf{K} > T_K$ and $\mathbf{H} < T_H$) and the nose bridge to be a saddle region ($\mathbf{K} < T_K$ and $\mathbf{H} > T_H$), where $T_K = 0.0025$ and $T_H = 0.00005$. Since there may be a number of pit regions, a systematic way is needed to find those corresponding to the eye cavities. First, small pit regions (< 80 points) are removed.
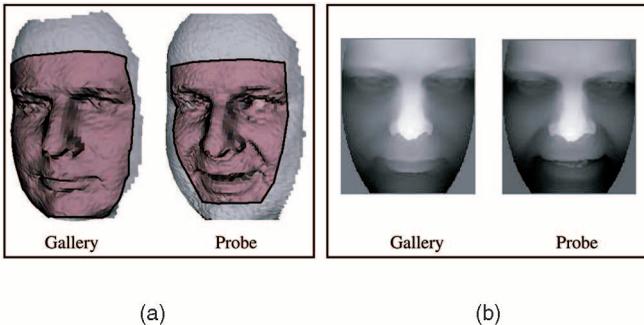

(a)   (b)

Fig. 2. Example of the large frontal face regions used with the baseline algorithms. For the ICP-baseline, note that the probe face is intentionally smaller than the gallery face, to ensure that the all the probe surface has some corresponding part on the gallery surface.

Second, a pair of regions that has similar average value in both Y and Z is found. Third, if there are still multiple candidate regions, the ones with higher Y values are chosen. The nose tip is found next. Starting between the eye landmark points, the search traverses down looking for the peak region with the largest difference in Z value from the center of the pit regions. Last, the area located between the two eye cavities is searched for a saddle region corresponding to the nose bridge.

3.3 Extracting Gallery/Probe Surface Patches

The pose is standardized in order to help make probe surface extraction more accurate. Pose correction is performed by aligning an input surface to a generic 3D face model. A circular region around the nose tip is extracted as a probe C, see Fig. 5b. Surface registration is performed between a probe C and a model surface. One reason to use a probe C rather than a whole facial region is to improve the registration in the presence of hair obscuring part(s) of the face. The input data points are then transformed using this registration.

For ICP-based matching, the probe surface should be a subset of the gallery surface, see Fig. 5a. For a probe, three different surfaces are extracted around the nose peak. These probe surfaces are extracted using predefined functions that are located on each face by the automatically-found feature points. For example, probe surface N is defined by a rectangle (labeled as 1 in Fig. 5c) based on the automatically-found nose tip and eye pit landmarks, with parts cut out based on four predefined ellipse regions (labeled as 2, 3, 4, and 5). Each of the elements is defined by constant offset values from the centroids of the facial landmark regions. Considering several different probe surfaces provides a better chance to select the best match among them. For instance, the probe N excludes the nostril portion of the nose while the probe I contains more of the forehead thus capturing more profile information of the nose. The definitions of these three probe surfaces were determined a priori based on considering results of earlier work [10].

Our curvature-based detection of facial landmark regions is fully automatic and has been evaluated on 4,485 3D face images of 546 people with a variety of facial expressions. The accuracy of the facial feature finding method is measured based on the degree of inclusion of the nose area in the probe C. The landmarks (eye cavities, nose tip, and nose bridge) were successfully found in 99.4 percent of the images (4,458 of 4,485). In those cases where the landmark points are not found correctly, a recognition error is almost certain to result.

TABLE 1
Description of the Data Set Used in This Study

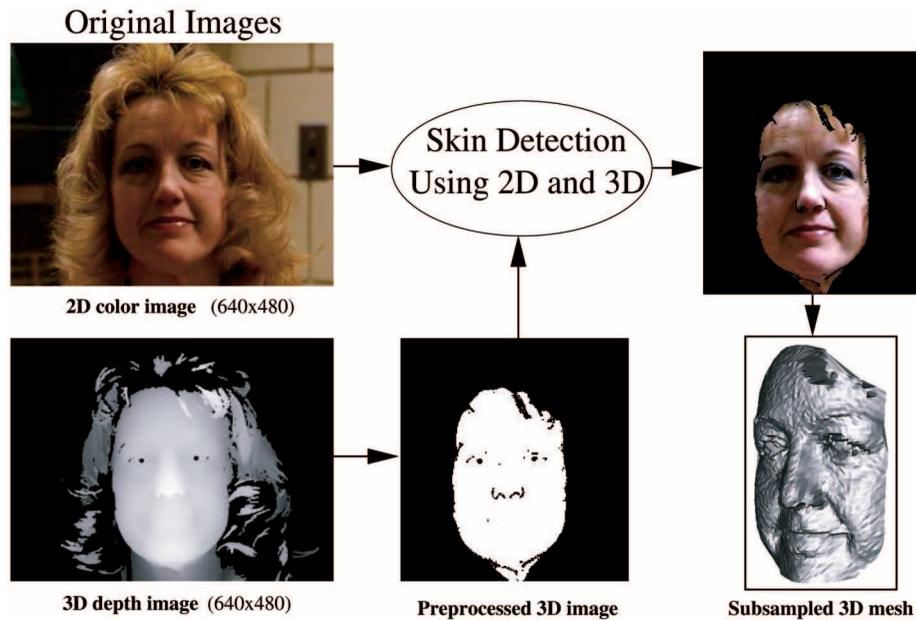| neutral expression	gallery	1	2	3	4	5	6	7	8	9
No. of subjects	*449*	449	390	336	286	243	205	172	145	123
Avg. days time-lapse	-	37	65	99	135	168	202	236	275	301
non-neutral	gallery	1	2	3	4	5	6	7	8	–
No. of subjects	*355*	355	321	266	209	168	125	91	55	–
Avg. days time-lapse	-	109	121	142	176	207	237	260	267	–

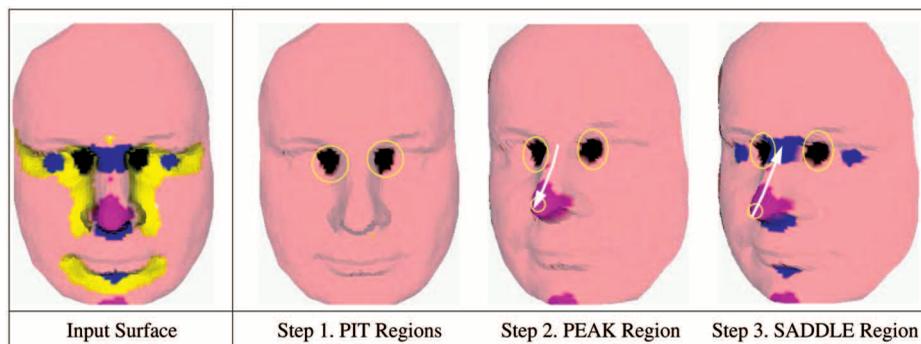Fig. 3. Illustration of steps in face region extraction.

Fig. 4 Steps involved in the facial landmark region detection process.

4 EXPERIMENTAL RESULTS

Table 2 gives rank-one recognition rates for the individual probe surfaces. As described earlier, the ICP baseline that matches a larger frontal surface achieves 91.0 percent rank-one recognition in matching neutral expression probe shapes to neutral expression gallery shapes. Interestingly, each of the three nose region surfaces individually achieves 95-96 percent rank-one recognition in neutral expression matching. The fact that *using less of the face can result in more accurate recognition* may at first seem contradictory. However, even if a subject is asked to make a neutral expression at two different times, the 3D face shape will still be different by some amount. Also, difficulties with hair over the forehead, or with noise around the regions of the eyes, are more likely with the larger frontal face region. Our result suggests that such "accidental" sources of variation are much less of a problem for the nose region than for larger face regions.

In the case of expression variation, the ICP baseline using the frontal face resulted in 61.5 percent rank-one recognition. As shown in Table 2, an individual nose region surface such as probe N achieves nearly 84 percent. Probe C has lower performance, possibly because it contains points in regions where more frequent deformation was observed. We next consider recognition from combining the results obtained from multiple nose region surfaces.

4.1 Performance Using Two Surfaces for a Probe

We considered three rules for combining similarity measurements from multiple probe surfaces: sum, product, and minimum. All three showed similar performance when matching a neutral expression probe to a neutral expression gallery: 96.59 percent average for product, 96.57 percent for sum, and 96.5 percent for minimum. However, in the presence of expression variation the product and sum achieved 87.1 percent and 86.8 percent, whereas the minimum rule achieved only 82.9 percent. Thus, we selected
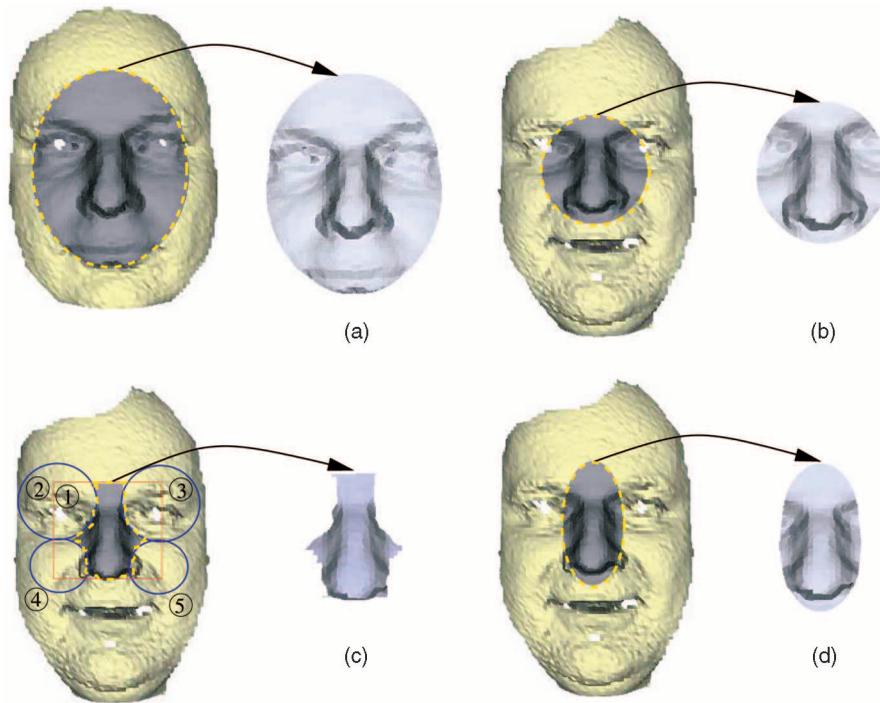
Fig. 5. Matching surface extraction for a gallery and three *probe surfaces*. (a) A gallery surface, (b) probe C (probe surface in general center face area, (c) probe N (probe surface in a nose region), and (d) probe I (probe surface in an interior nose region).

TABLE 2
Rank-One Recognition Rates for Individual Probe Surfaces

Neutral expression probe matched to neutral expression gallery										
	# 1	# 2	# 3	# 4	# 5	# 6	# 7	# 8	# 9	Mean
probe C	95.6	94.6	94.9	96.2	95.5	95.6	95.4	95.9	99.2	95.9
probe N	94.0	94.4	93.8	94.4	95.5	95.1	95.9	95.2	100.00	95.4
probe I	96.2	95.6	95.5	95.8	96.7	95.6	97.1	96.6	100.00	**96.6**

Non-neutral-expression probe to neutral-expression gallery									
	# 1	# 2	# 3	# 4	# 5	# 6	# 7	# 8	Mean
probe C	63.7	75.1	75.2	78.5	79.8	78.4	76.9	89.1	77.1
probe N	76.1	81.9	80.1	81.3	86.9	87.2	85.7	89.1	**83.5**
probe I	77.2	81.9	79.7	81.8	83.9	86.4	83.5	87.3	82.7

the product rule, although its results are not significantly different from using the sum rule.

The results for using different combinations of the individual probe surfaces are shown in Table 3. When matching neutral expression to neutral expression, the information in the different nose region surfaces is somewhat redundant. Comparing Table 3 to Table 2, the performance improvement is less than one percent. However, when there is expression change between the probe and the gallery, combining results from multiple nose region surfaces has a larger effect. In this case, the best individual probe surface resulted in 83.5 percent rank-one recognition and the best pair of surfaces resulted in 87.1 percent. Interestingly, while the combination of three surfaces improved slightly over the best combination of two surfaces in the case of matching neutral expressions, the combination of three did not do as well as the best combination of two in the case of matching varying expressions. In the end, the best overall performance comes from the combination of probe surfaces N and I. One surprising element of this work is that we can achieve such good performance using only a small portion of the face surface. However, there still exists an approximate 10 percent performance degradation, from roughly 97 to 87 percent, in going from matching neutral expressions to matching varying expressions.

The Receiver Operating Characteristic (ROC) curve in Fig. 6 reports results for a verification scenario. The equal-error rate (EER) is the ROC point at which the false accept rate is equal to the false reject rate. The EER for our approach goes from approximately 0.12 for neutral expressions to approximately 0.23 for varying expressions. The EER for the ICP baseline shows a much greater performance degradation in going from all neutral expressions to varying facial expression. This indicates that our new algorithm is effective in closing part of performance gap that arises in handling varying facial expressions.

A verification scenario implies 1-to-1 matching of 3D shapes, whereas a recognition scenario implies matching one probe against a potentially large gallery. ICP-based matching of face shapes can be computationally expensive. Our current algorithm takes approximately one second to match one probe shape against one gallery shape. Techniques to speed up 3D shape matching in face recognition are a topic of current research [18], [19].

The results in Fig. 7 show the effect on recognition rate of varying the number of enrolled subjects. We begin with probe set #1, randomly select one half of probe set #1 to generate a reduced-size probe set, and do this multiple times. To account for variations in subject pool, the performance shown for each reduced data set size is the average of 10 randomly selected subsets of that size. There is a

TABLE 3
Rank-One Recognition Rates Using Multiple Probe Surfaces

Neutral-expression probe to neutral-expression gallery										
	# 1	# 2	# 3	# 4	# 5	# 6	# 7	# 8	# 9	Mean
probes $C+N$	96.4	96.9	96.1	97.2	96.3	96.1	97.1	95.9	99.2	96.8
probes $C+I$	96.9	96.4	95.8	96.9	96.7	96.1	95.9	97.2	100.0	96.9
probes $N+I$	96.7	96.7	94.9	95.8	96.7	95.6	97.1	95.9	100.0	96.6
probes $C+N+I$	96.9	96.7	95.8	96.9	97.1	95.6	96.5	97.2	100.0	97.1

Non-neutral-expression probe to neutral-expression gallery									
	# 1	# 2	# 3	# 4	# 5	# 6	# 7	# 8	Mean
probes $C+N$	75.5	84.4	81.6	85.2	86.3	90.4	84.6	92.7	85.1
probes $C+I$	74.4	81.0	79.3	84.2	84.5	88.0	83.5	89.1	83.0
probes $N+I$	80.6	85.4	81.2	87.1	87.5	91.2	91.2	92.7	87.1
probes $C+N+I$	79.2	85.7	80.8	86.1	88.7	88.0	85.7	94.6	86.1

trend toward higher observed rank-one recognition rate with smaller gallery size. This effect is much more prominent when expressions are varied. However, the degree of decrease in recognition rate that accompanies a doubling in gallery set size is much less here for 3D than has been reported by others for 2D [20].

5 SUMMARY AND DISCUSSION

We consider the issue of facial expression variation in 3D face recognition. Results of our new approach are compared to results from PCA and ICP-based approaches similar to previous work. Our new approach uses multiple overlapping surfaces from the nose area since this area appears to have relatively low shape variation across a range of expressions. Surface patches are automatically extracted from a curvature-based segmentation of the face. We consider using as many as three overlapping probe surface patches, but find that three does not improve performance over using two. Our approach substantially outperforms the ICP baseline that uses a frontal face region and manually identified landmark points. However, there is more to be done to solve the problem of handling expression variation, as there is about a 10 percent drop in rank-one recognition rate when going from matching neutral expressions to matching varying expressions.
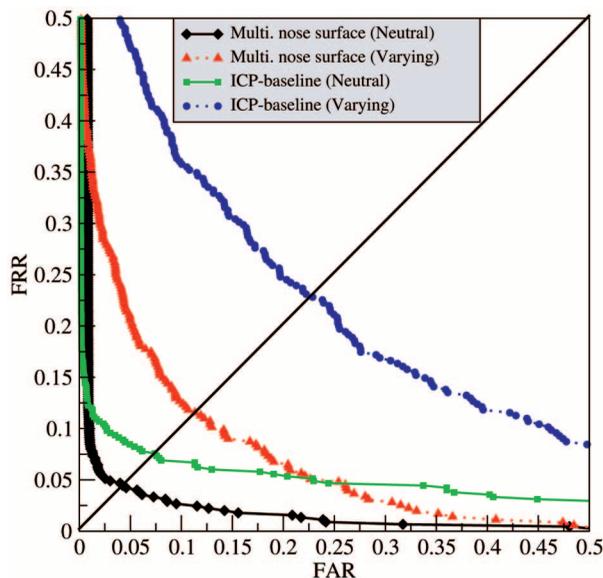

Fig. 6. ROC performance on neutral and non-neutral expression probe sets.

One possible means to better performance is to use additional probe regions. For example, surface patches from the temples and/or from the chin may carry useful information about face shape and size. Algorithms to use such larger collections of surface patches will need to deal with missing patches, and make comparisons across probes that may use different numbers of patches in matching. The work of Cook et al. [3] may be relevant in this regard. They experimented with an approach to 3D face recognition that uses ICP to register the surfaces, then samples the distance between the registered surfaces at a number of points and models the intra versus interperson distribution of such feature vectors. It may be possible to adapt this approach to deal with expression variation, either by registering parts of the face surface individually, or by detecting elements of interperson variation caused by change in facial expression.

There has been substantial work in dealing with expression variation in 2D face recognition. Yacoob et al. suggested "that there is need to incorporate dynamic analysis of facial expressions in future face recognition systems to better recognize faces" [21]. This seems promising for future work, but sensors for 3D face imaging are currently not as mature as 2D camera technology [17]. Martinez has also worked on 2D face recognition in the context of facial expressions and noted that "different facial expressions influence different parts of the face more than others" [22]. He developed a strategy for "giving more importance to the results obtained from those local areas that are less affected by the current displayed emotion" [22]. This general motivation is similar to that in our work. Also, Heisele has done work looking at "components" of the face [23]. He experimented with 14 local regions, or components, of 2D face appearance using 3D morphable models, and presented "a method for automatically learning a set of discriminatory facial components" in this context [23]. The automatic learning of useful local regions of the 3D face shape is an interesting topic for future research.
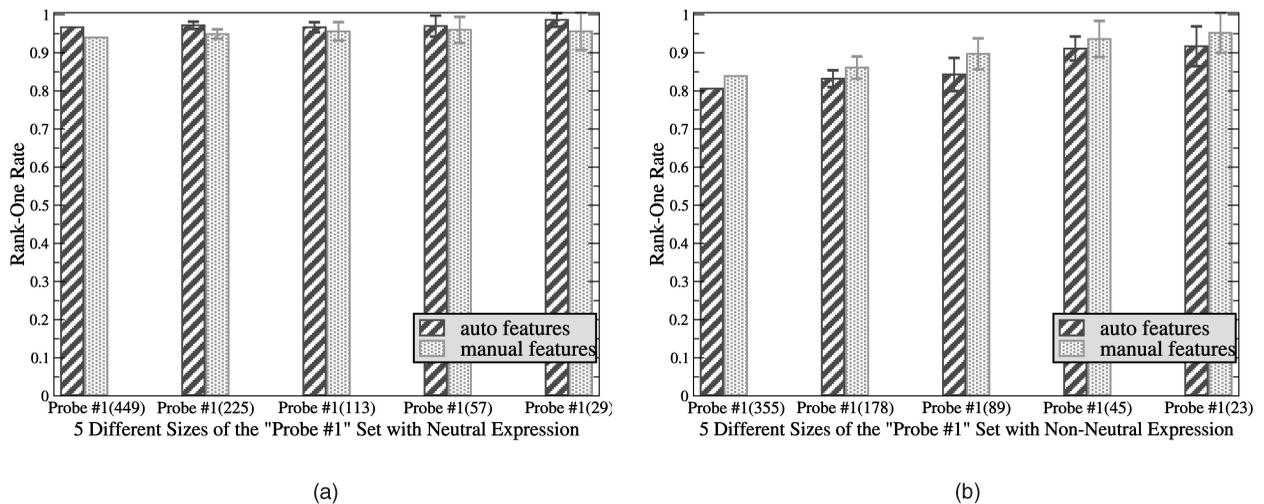
(a)

(b)

Fig. 7. Rank-one recognition rate with varying sata set size. (a) Neutral expression probes and (b) varying expression probes.

REFERENCES

[1] C. Hesher, A. Srivastava, and G. Erlebacher, "A Novel Technique for Face Recognition Using Range Imaging," *Proc. Int'l Symp. Signal Processing and Its Applications,* pp. 201-204, 2003.

[2] G. Medioni and R. Waupotitsch, "Face Modeling and Recognition in 3-D," *Proc. IEEE Int'l Workshop Analysis and Modeling of Faces and Gestures,* pp. 232-233, Oct. 2003.

[3] J. Cook, V. Chandran, S. Sridharan, and C. Fookes, "Face Recognition from 3D Data Using Iterative Closest Point Algorithm and Gaussian Mixture Models," *Proc. Second Int'l Symp. 3D Data Processing, Visualization, and Transmission,* pp. 502-509, 2004.

[4] K.I. Chang, K.W. Bowyer, and P.J. Flynn, "Adaptive Rigid Multi-Region Selection for Handling Expression Variation in 3D Face Recognition," *Proc. IEEE Workshop Face Recognition Grand Challenge Experiments,* June 2005.

[5] M. Husken, M. Brauckmann, S. Gehlen, and C. von der Malsburg, "Strategies and Benefits of Fusion of 2D and 3D Face Recognition," *Proc. IEEE Workshop Face Recognition Grand Challenge Experiments,* June 2005.

[6] X. Lu and A.K. Jain, "Deformation Analysis for 3D Face Matching," *Proc. Seventh IEEE Workshop Applications of Computer Vision,* pp. 99-104, 2005.

[7] T. Maurer, D. Guigonis, and I. Maslov et al., "Performance of Geometrix Activeid 3D Face Recognition Engine on the FRGC Data," *Proc. IEEE Workshop Face Recognition Grand Challenge Experiments,* June 2005.

[8] G. Passalis, I.A. Kakadiaris, T. Theoharis, G. Toderici, and N. Murtuza, "Evaluation of 3D Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach," *Proc. IEEE Workshop Face Recognition Grand Challenge Experiments,* June 2005.

[9] C. Chua, F. Han, and Y. Ho, "3D Human Face Recognition Using Point Signature," *Proc. Int'l Conf. Automatic Face and Gesture Recognition,* pp. 233-238, 2000.

[10] K. Chang, K. Bowyer, and P. Flynn, "Effects on Facial Expression in 3D Face Recognition," *Proc. SPIE Conf. Biometric Technology for Human Identification,* pp. 132-143, Apr. 2005.

[11] A.M. Bronstein, M.M. Bronstein, and R. Kimmel, "Three-Dimensional Face Recognition," *Int'l J. Computer Vision,* vol. 64, pp. 5-30, 2005.

[12] A.M. Bronstein, M.M. Bronstein, and R. Kimmel, "Generalized Multi-dimensional Scaling: A Framework for Isometry-Invariant Partial Surface Matching," *Proc. Nat'l Academy of Science,* pp. 1168-1172, 2006.

[13] K.I. Chang, K.W. Bowyer, and P.J. Flynn, "An Evaluation of Multi-Modal 2D + 3D Face Biometrics," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, pp. 619-624, 2005.

[14] F. Tsalakanidou, D. Tzovaras, and M. Strintzis, "Use of Depth and Colour Eigenfaces for Face Recognition," *Pattern Recognition Letters,* pp. 1427-1435, 2003.

[15] X. Lu, A.K. Jain, and D. Colbry, "Matching 2.5D Face Scans to 3D Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, pp. 31-43, 2006.

[16] B. Gokberk, A. Ali Salah, and L. Akarun, "Rank-Based Decision Fusion for 3D Shape-Based Face Recognition," *Proc. Fifth Int'l Conf. Audio and Video-Based Biometric Person Authentication,* pp. 1019-1028, July 2005.

[17] K. Bowyer, K. Chang, and P. Flynn, "A Survey of Approaches and Challenges in 3D and Multi-Modal 3D+2D Face Recognition," *Computer Vision and Image Understanding,* vol. 101, no. 1, pp. 1-15, 2006.

[18] P. Yan and K.W. Bowyer, "A Fast Algorithm for ICP-Based 3D Shape Biometrics," *Proc. IEEE Workshop Automatic Identification Advanced Technologies,* Oct. 2005.

[19] M.L. Koudelka, M.W. Koch, and T.D. Russ, "A Prescreener for 3D Face Recognition Using Radial Symmetry and the Hausdorff Fraction," *Proc. IEEE Workshop Face Recognition Grand Challenge Experiments,* June 2005.

[20] J. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, and M. Bone, "Facial Recognition Vendor Test 2002: Evaluation Report," http://www.frvt2002.org/FRVT2002/documents.htm, 2002.

[21] Y. Yacoob, H.-M. Lam, and L. Davis, "Recognizing Faces Showing Expressions," *Proc. Int'l Workshop Automatic Face and Gesture Recognition,* pp. 278-283, 1995.

[22] A. Martinez, "Recognizing Imprecisely Localized, Partially Occluded, and Expression Variant Faces from a Single Sample per Class," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, pp. 748-763, June 2002.

[23] B. Heisele and T. Koshizen, "Components for Face Recognition," *Proc. Sixth IEEE Int'l Conf. Face and Gesture Recognition,* pp. 153-158, 2004.

[24] P.J. Phillips et al., "Overview of the Face Recognition Grand Challenge," *Computer Vision and Pattern Recognition,* pp. I:947-954, June 2005.