3D face analysis for demographic biometrics

Ryan Tokola, Aravind Mikkilineni, and Christopher Boehnen
Oak Ridge Nationall Laboratory
tokolara, mikkilinenak, boehnencb@ornl.gov

Abstract

Despite being increasingly easy to acquire, 3D data is rarely used for face-based biometrics applications beyond identification. Recent work in image-based demographic biometrics has enjoyed much success, but these approaches suffer from the well-known limitations of 2D representations, particularly variations in illumination, texture, and pose, as well as a fundamental inability to describe 3D shape.

This paper shows that simple 3D shape features in a face-based coordinate system are capable of representing many biometric attributes without problem-specific models or specialized domain knowledge. The same feature vector achieves impressive results for problems as diverse as age estimation, gender classification, and race classification.

1. Introduction

The last twenty years have seen significant advances in face recognition as well as increasing interest in less specific forms of face-based biometrics. In many ways it is more difficult to infer information about a person (what we are calling *demographic* biometrics) than to determine their identity. Investigations into age estimation and gender and race classification have been fruitful, but the proposed methods are very specific to their particular domain. For example, many age estimation algorithms rely on wrinkle detection [10] and gender classifiers may use distances between specially designated landmarks [6].

With the increasing availability of 3D sensors it is natural to consider 3D shape as an important component of facial biometrics. 3D data not only captures information that is unavailable in 2D images, but it is also insensitive to many of the factors that make image-based biometrics difficult, such as pose, illumination, and color. 3D data has been successfully used for face recognition [3][5], but very few approaches to demographic biometrics have taken advantage of this rich set of features. This may be because 3D scans are rarely fine-grained enough to capture wrinkles for age estimation, and cannot represent makeup for gender recognition or skin pigmentation for race classification. This paper shows, however, that color and texture information are not required for good performance in these tasks.

The purpose of this paper is to show that 3D data can be useful for biometrics applications beyond recognition. We show that a very simple set of 3D-based features can be used to learn much more than a person's identity. The features used in this paper are very straightforward—3D points are projected into a face-based coordinate system [3] and then transformed with Principal Component Analysis (PCA) in the spirit of eigenfaces [13].

Unlike many approaches to biometrics these features are not tuned to any specific experiment and no domain-specific knowledge is used. Despite being very simple, the 3D features have surprising discriminative power. Experiments in Section 5 show that *the exact same feature vector* can be fed into task-specific classifiers to achieve an age estimation Mean Absolute Error (MAE) of 4.9 years, a gender classification accuracy of 88.8%, and a race classification accuracy of 99.9%.

A face-based coordinate system is also exceptionally well-suited to intuitive analysis and visualization. As will be explained in Section 3, the coordinate system describes each point on a face in relation to a reference face. Figure 1 shows the effects of altering face shapes to more clearly illustrate differences between groups. The top row shows the differences between male and female Caucasians. The center face is the overall average Caucasian face, and the faces to the left and right of the center show the average female and male faces, respectively. Notice that it is difficult to identify the differences between the two groups using the

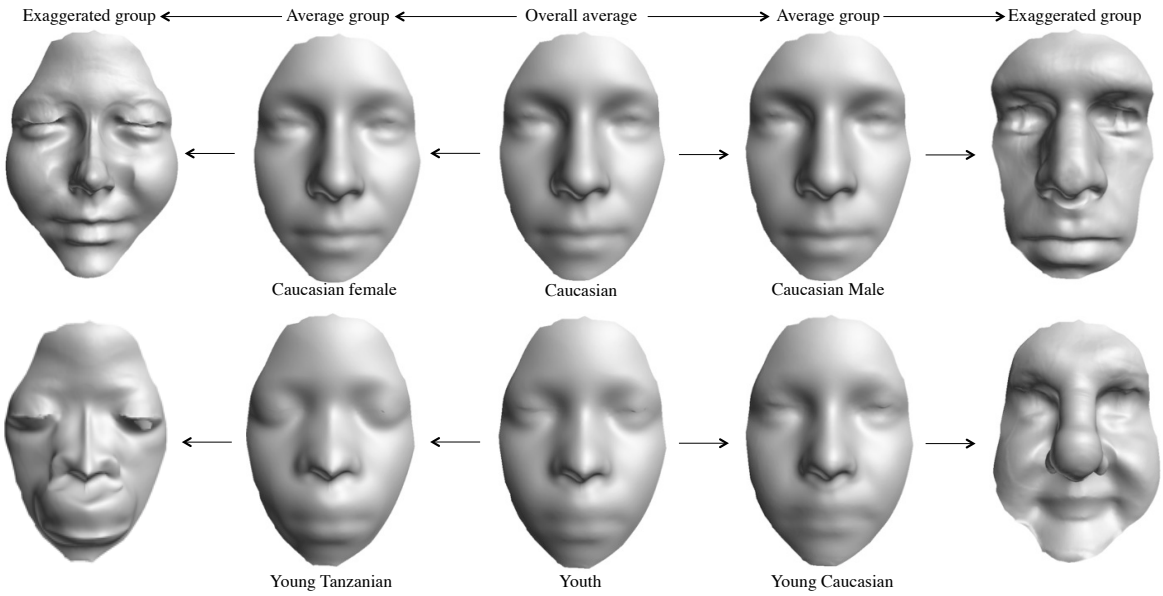


Figure 1. Exploring differences within two pairs of demographic groups using exaggeration. In all cases, calculations were performed in a face-based coordinate system and the results were backprojected to a 3D mesh. The demographic pairs are Caucasian male/Caucasian female and young Tanzanian/young Caucasian. Overall average faces are shown in the center, group averages are shown next, and exaggerated faces are shown at the ends. Note that the exaggerated faces allow for an easier understanding of the differences between average group faces.

average faces. The differences were exaggerated in the face-based coordinate system, resulting in the faces at the ends of the row. Here, the differences between male and female faces become much more obvious. We can see, for example, that female Caucasians tend to have more pronounced cheekbones, and male Caucasians tend to have more prominent eyebrows. The bottom row of Figure 1 is structured similarly, except the two groups are young Caucasians and young Tanzanians. Gender labels for the Tanzanian subjects in the Spritz dataset (see Section 4.2) are not available, so the gender differences between Tanzanian faces cannot be explored.

This paper does not claim to have the best possible performance in any individual demographic biometrics application. Instead, it demonstrates that a modest set of 3D features contains a promising amount of information. It is reasonable to assume that performance on any particular classification task could be improved by incorporating 2D image data, problem-specific models, or more sophisticated use of the 3D data. We hope to convince the reader that 3D data in a face-based coordinate system is a good starting point.

2. Related works

The literature related to demographic biometrics is vast, encompassing 2D and 3D approaches to face recognition, age estimation, and gender and race recognition. It is not possible in this paper to present complete introductions to all of the topics. Instead, we refer the reader to a few comprehensive surveys for a broad understanding of the issues in each application. Zhang and Zhang [17] survey face recognition, with an emphasis on 2D methods. Bowyer et al. [4] survey face recognition using 2D and 3D data. Fu et al. [7] survey race classification. Guo surveys gender classification and age estimation [8].

This work is closely related to one of the seminal approaches to face recognition: the eigenfaces of Turk and Pentland [13]. In both cases classification is performed using PCA-based projections.

In the biometrics community, 3D data has most commonly been used for face recognition. Blanz and Vetter [2] use a 3D morphable model for face recognition, but their recognition strategy does not exclusively use 3D data. Instead, they use a combination of shape and texture coefficients from the morphable model. Russ et al. [11] align 3D query faces with a reference face, project the points of the reference face onto the query face, and apply PCA to the resulting 3D points. This is the inspiration for this paper, the primary difference being that we use the 3D signatures of Boehnen et al. [3] instead of 3D points.

A method called 3D eigenfaces was proposed by Xu et al. [15], which first aligns faces with an average face and then uses a hierarchical procedure to fit a regular mesh to the query face. PCA is applied to the z-coordinate of the fitted mesh.

Drira et al. [5] are a good example of a more contemporary approach to 3D face recognition. They apply elastic shape analysis to 3D curves that radially emanate from the nose tip. They show strong performance on data with significant occlusions and variation in expression and pose.

There are surprisingly few works applying 3D face shape to biometric attributes other than identity. Hu et al. [9] estimate gender using 3D face information, but their features are limited to the location of a handful of facial landmarks.

Xia et al. [14] perform age estimation from 3D data using an approach that is similar to [5] in that they use features derived from curves that radiate from the tip of the nose. Unfortunately, the dataset used for their experiments has a poor distribution across ages, and so is not well suited to age estimation. [14] report a MAE of 3.15 years on the dataset, but it should be noted that simply guessing that all subjects are 21 years old will result in a MAE of only 3.7 years.

3. Correspondence Vectors (CVs)

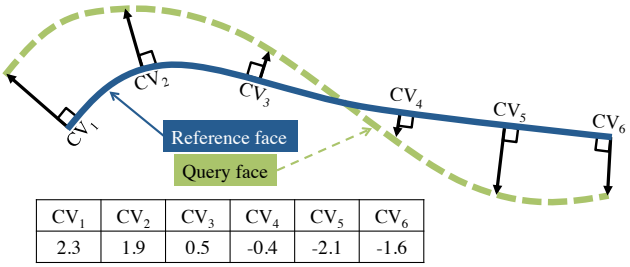
This paper demonstrates that simple 3D features can be applied to several forms of demographic biometrics and lend themselves to an intuitive understanding of how these biometrics are reflected in face shape. We avoid the complex geometric reasoning of [5] and [14], but are interested in features with more expressive potential than a few manually defined landmarks [9]. We have found the 3D signatures from Boehnen et al. [3] to be a fruitful compromise. The term "3D signature" is intended to convey a sense of individuality and is used in the context of face recognition. We will use "Correspondence Vector (CV)" instead, emphasizing that the feature is a measure of the *correspondence* between two faces.

A CV measures the distance between a query face and a reference face, mapping points from a global coordinate system to face-based coordinates. CVs enjoy three advantages over features in a global coordinate system:

- The CV transforms a 3D vector (x, y, z) into a 1D vector (d), resulting in decreased computational complexity and a simpler PCA calculation.

- CVs are easier to interpret. For example, if an element of a CV has a value of -0.2, then the point is located 0.2 units behind the reference face. If the point is on a cheek, it could contribute to a gaunt appearance. It is much more difficult to associate an absolute 3D coordinate, say (18.2, 29.0, 11.7), with any such attributes.

- CVs are easier to modify in meaningful ways. For example, a 3D face can be warped to be "more male" by taking the CV of the face and adding the difference between an average male face and an overall average face. This is shown in Figure 1.



CV_1	CV_2	CV_3	CV_4	CV_5	CV_6
2.3	1.9	0.5	-0.4	-2.1	-1.6

Figure 2. Schematic illustration of a CV with six elements.

3.1. Mesh alignment

Before measuring the distance between a query face mesh and a reference face, it is first necessary to align them. The Iterative Closest Point (ICP) algorithm [1] is a commonly used approach to aligning point clouds. Given a rough initial alignment, the algorithm moves the query face in 3D space to minimize the distance between the two faces. After registration, every vertex in the reference face is projected along the surface normal and the signed distance to the query face is measured. This is illustrated in Figure 2. [11] and [3] also discuss an alternative nearest neighbor approach, but the distance along the normal is particularly well suited to a simple analysis of 3D shape. Please refer to [3] for a more thorough discussion of reference face alignment and the generation of CVs.

3.2. Principal components of CVs

In a manner similar to eigenfaces [13], we obtain image features by applying PCA to CVs. As 1D vectors CVs naturally lend themselves to PCA. The reference face used in this work contains 15,130 vertices, which is a large number of features. Fortunately, there is much redundant information in a full CV, and the dimensionality can be dramatically reduced without sacrificing valuable relationships within the data. As shown in Figure 3 only 100 Principal Components (PCs) are required to explain 93% of the variance in the Weinberg dataset, which is described in Section 4.1.

Figure 4 shows a graphical representation of the first twelve PCs from the Weinberg dataset. The elements of each PC are color-mapped with red indicating more positive values and blue indicating more negative values. Since the original CVs can be interpreted as a set of distances from a standard reference face, the PCs have an intuitive meaning in that the value of the PC at a specific point indicates whether the point lies in front of or behind the reference face. Regions where the PC is positive (red) valued will be in front of the reference face whereas regions where the PC is negative (blue) tend to be behind the reference face. A yellow color indicates no change from the reference face. Consider the first PC shown in Figure 4. A negative value

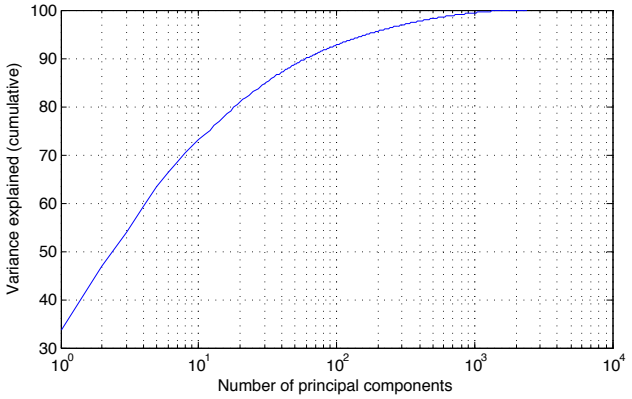


Figure 3. Cumulative percent variance captured by the first 100 principal components from the Weinberg dataset (full face).

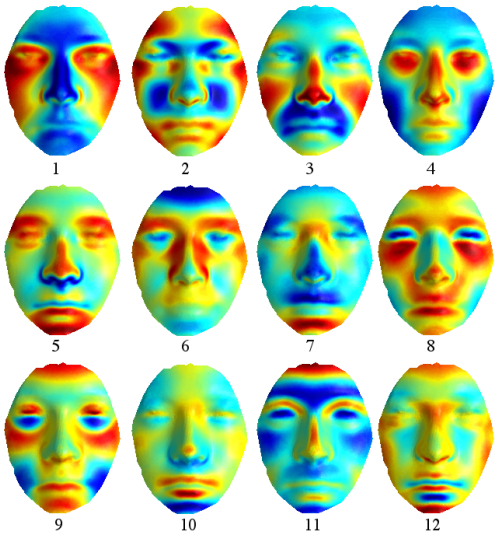


Figure 4. First twelve PCs from the Weinberg dataset using the full reference face.

for this component corresponds to a face in which the nose, forehead, and chin protrude more than they do with an average face. Similarly a positive value for the 11th component corresponds to a face that has more prominent eyebrows.

Figure 5 shows a graphical representation of the first twelve PCs from the Spritz dataset, which is described in Section 4.2. Note that the first PC is surrounded by a positive (red) band. This band is strongly related to overall scale, which is reasonable, considering that the Spritz dataset is composed of children's faces, which exhibit greater variation in scale.

3.3. Reverse mapping CVs to faces

One of the most interesting characteristics of CVs is that it is simple to recreate a face mesh given a CV. This is very

Figure 5. First twelve PCs from the Spritz dataset using the full reference face.

different from other biometrics algorithms that rely upon statistics or complex curve analysis. To generate a face mesh, it is only necessary to shift each point of the reference face the distance given by the CV in the direction of the surface normal. The ability to map CVs to 3D faces enables some interesting visualization and analysis as discussed in Sections 1, 5.2, and 5.3.

3.4. Reference Face Generation

For CVs to accurately describe a 3D shape, it is important for the reference face to be smooth. This ensures that the surface normals represent the overall shape of the reference face. It must also be constructed such that the ICP algorithm can provide the best possible alignment of faces. Reference faces are created using an iterated procedure. First several face meshes are averaged to make an initial reference face. Then CVs are calculated for a training set given the reference face. The CVs are averaged, and the resulting mean CV is reverse mapped to create a new reference face. This is repeated until convergence. Please refer to [3] for more details.

4. Datasets

All 3D data used in this paper is provided by the FaceBase Consortium, which is an NIH-supported initiative. The data consists of two distinct datasets. One was contributed by Dr. Seth Weinberg and contains 3D face meshes of primarily Caucasian subjects. This will be referred to as the Weinberg dataset. The other was contributed by Dr. Richard Spritz and contains 3D face meshes of Tanzanian subjects. This will be referred to as the Spritz dataset. Both

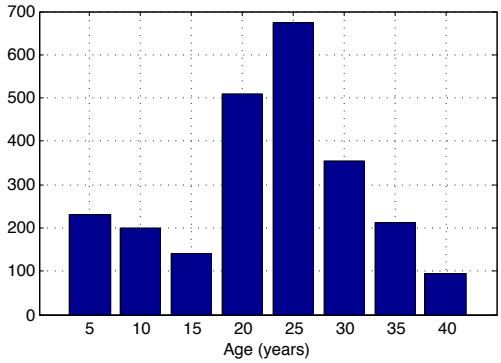


Figure 6. Age distribution of the Weinberg dataset.

datasets are available at no cost through the NIH.

4.1. Weinberg dataset

The Weinberg dataset contains 2454 3D face meshes in Wavefront (.obj) format. Metadata includes the age, race, and gender of each subject. Alignment failed for 39 of the meshes, so the remaining 2415 were used for experiments. Each mesh covers at least half the skull from the tip of the nose back to the ears. Subjects range in age from 3 to 40 with the distribution shown in Figure 6. 952 subjects are male and 1502 subjects are female. The subjects are mostly Caucasian.

Application of the CV generation method to faces in the Weinberg dataset requires some preprocessing to obtain correct alignment. The faces were oriented such that they were facing in the same direction as the reference face with the tip of the nose placed at the origin.

4.2. Spritz dataset

The Spritz dataset contains 3605 3D face meshes in Wavefront (.obj) format. No demographic data (other than the fact that the country of residence is the United Republic of Tanzania) is provided. The subjects appear to range in age from 3 to 17 years old, with most subjects between 8-12 years.

5. Experiments

In this section three sets of experiments are used to demonstrate that CVs perform well on a variety of demographic biometrics tasks despite a complete lack of specialized features, models, or domain knowledge.

Because of their general nature it cannot be expected that CVs should outperform the most current, specialized, and complex biometrics algorithms. The purpose of these experiments is to show that 3D facial structure is useful for purposes other than face recognition and to encourage its use in future research. We hope that CVs will be useful as a baseline for more sophisticated algorithms.

On a desktop computer all preprocessing (ICP alignment, CV extraction, and PCA projection) takes approximately 8 seconds per face. However, once this processing has been done the feature vector can be evaluated extremely rapidly. One of the advantages of our general-purpose feature is that there is no need to extract new task-specific features for each classification problem.

5.1. Age estimation

The first set of experiments explores the performance of CVs on the task of age estimation. The Weinberg dataset was randomly partitioned into a training set of 1932 samples and a testing set of 483 samples. A Support Vector Regressor (SVR) was learned using the training set. Both Radial Basis Function (RBF) kernels and linear kernels were used for the SVR. Values for C (and γ in the case of a RBF kernel) were selected using five-fold cross-validation on the training set.

Results are shown in Table 1 with a varying number of PCs. The value in the row labeled "number PCs" indicates the number of PCs that were used in the experiment.

The mean age of the Weinberg dataset is 22.3 years. A classifier that guesses the mean age will result in a Mean Squared Error (MSE) of 82.3 years, or a MAE of 9.1 years.

The best performing classifier uses 200 PCs and a SVR with a RBF kernel. It has a MSE of 23.5 years, or a MAE of 4.8 years. Note that the performance of classifiers decreases slightly with more than 200 PCs. This indicates that all relevant age-based information is captured in the first 200 PCs. This is consistent with the curve seen in Figure 3, which shows that 200 PCs capture approximately 96% of the variance in the dataset.

Age estimation results are shown partitioned by ground truth age in Figure 7. The performance of the 3D features is compared to a classifier that always guesses the average age of the test set, which is 22.3 years. It can be seen that the 3D features outperform the average age-based classifier, except in the immediate vicinity of the mean age. The 3D features show an unexpected decrease in performance in the 18-22 year age bin, but it is only 0.6 years worse than the 23-27 year age bin.

5.2. Gender classification

The second set of experiments applies CVs to gender classification. The experimental setup is the same as in Section 5.1, except Support Vector Machine (SVM) classifiers are learned instead of SVRs. Because the genders are not equally distributed in the dataset, a classifier that always guesses female will have an accuracy of 61.2%.

Results are shown in Table 1. Classifiers have a peak classification rate of 89.2%. It is somewhat surprising to note that this result is from an SVM with a linear kernel, and not an RBF kernel. As is the case with age estimation,

Age estimation performance (Mean Absolute Error, smaller is better)													
number PCs	1	2	3	4	5	10	20	50	100	200	300	400	500
RBF SVM	7.5	6.7	6.7	6.4	6.0	5.9	5.6	5.0	4.9	**4.8**	5.3	5.4	5.6
linear SVM	8.0	7.4	6.8	6.7	6.2	6.0	5.8	5.3	5.2	5.4	5.4	5.6	5.8

Gender classification rate (%, larger is better)													
number PCs	1	2	3	4	5	10	20	50	100	200	300	400	500
RBF SVM	78.7	80.1	81.8	81.6	81.2	86.1	86.5	87.2	88.8	88.6	88.0	86.7	87.2
linear SVM	78.7	80.7	79.7	79.9	81.0	86.3	85.1	87.4	88.4	**89.2**	87.2	87.2	85.5

Race classification rate (%, larger is better)													
number PCs	1	2	3	4	5	10	20	50	100	200	300	400	500
RBF SVM	93.1	92.9	94.6	99.0	99.4	99.6	99.8	**99.9**	**99.9**	99.0	98.9	99.1	99.1
linear SVM	93.1	93.1	94.8	98.8	99.0	99.1	99.6	**99.9**	99.6	99.6	99.5	99.1	99.4

Table 1: Experimental results on the Weinberg and Spritz datasets.

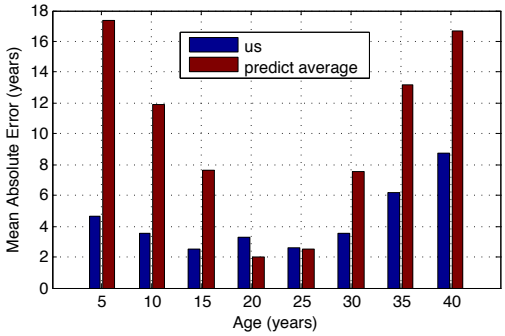


Figure 7. Age estimation performance. Results partitioned by age. A linear SVM was used for classification using 200 PCs.

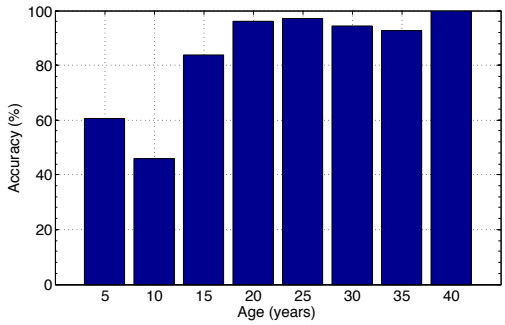


Figure 8. Gender classification performance. Results partitioned by age. An SVM with a RBF kernel was used for classification with 200 PCs.

the best performing classifier uses 200 PCs.

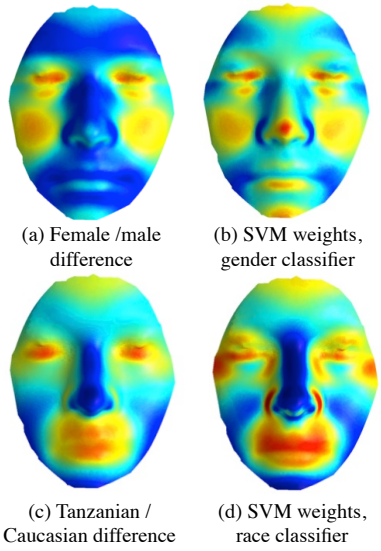
It is important to remember that the data was not curated specifically for this experiment. Many experiments in gender classification remove extra sources of variability by only considering adult subjects. This experiment, on the other hand, uses data from people as young as three years old. Figure 8 shows how gender classification performance varies with the subjects' age. Note that the performance is best among adults.

It has already been shown that a face-based coordinate system is convenient for visualization in terms of 3D meshes and heatmaps. In addition to visualizing CVs as heatmaps, it is also possible to visually analyze the weights of linear classifiers. Figure 9 compares the average CVs of demographic groups to the weights of classifiers used to identify them. Figure 9(a) shows the difference between the average female and the average male as a heatmap. (b) shows the weights of a linear female/male classifier. Notice the differences in the nose. The classifier responds more strongly (in a positive sense) to the protrusion of the tip of the nose, and more strongly (in a negative sense) to the width of the nose. From this we can see that the nose is among the most discriminative parts of the face for gender classification.

5.3. Race classification

The final set of experiments uses CVs for race classification among young subjects. The experimental setup is the same as in Section 5.2, except subjects from the Weinberg dataset are used as examples of the "Caucasian" class and the Spritz dataset is used as examples of the "Tanzanian" class. Only subjects 15 years old and younger were used from Weinberg dataset to approximately match the age distribution of the Spritz dataset.

3291 samples were randomly selected for testing and the remaining 822 samples were used for testing. As was the case for gender classification this is an especially challenging collection of data because the variation in face shape is due to age as well as race.



(a) Female /male difference

(b) SVM weights, gender classifier

(c) Tanzanian / Caucasian difference

(d) SVM weights, race classifier

Figure 9. Comparing the average difference between demographic groups to the weights of linear SVM classifiers. Values are shown as a heatmap. (a) Difference between average CVs of female and male Caucasian faces in the Weinberg dataset. (b) Weights of a linear gender classifier. (c) Difference between average Tanzian and Caucasian youths. (d) Weights of a linear race classifier.

Classifiers with an RBF kernel using either 50 or 100 PCs had a remarkable 99.9% race classification accuracy. As before, note the slight decrease in performance with additional PCs.

Figure 9(c) and (d) show the average CVs of young Tanzanians and the weights of a Tanzanian/Caucasian classifier, respectively. Notice that here the breadth of the nose is again a very important feature for discrimination.

6. Future work

CVs in a face-based coordinate system have been shown to be an effective and general means of addressing several biometrics problems. Just as image-based biometrics performance can be improved by combining PCA projections of images with other features [16] [12], CVs could contribute a robust description of 3D shape to an ensemble of more specialized features. We expect CVs to enhance the performance of models with problem-specific features.

The general and compact nature of CVs make them well-suited to large-scale face searches. It would be possible to pre-calculate the PCA projections of CVs for a large number of face scans and then apply a custom query extremely rapidly. A linear classifier could be constructed using labeled training data (say, Caucasian women over 40) and then applied to the entire database with a single matrix multiplication.

7. Conclusions

This paper has shown that extremely simple 3D features operating in a face-based coordinate system are general enough that they can be applied to several demographics biometrics problems, yet are powerful enough to provide impressive results for each problem. These features are easy to compute and lend themselves to intuitive and semantically meaningful interpretation.

References

[1] P. J. Besl and N. D. McKay. Method for registration of 3-D shapes. In *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics, 1992.

[2] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1063–1074, 2003.

[3] C. Boehnen, T. Peters, and P. J. Flynn. 3D signatures for fast 3D face recognition. In *Advances in Biometrics*, pages 12–21. Springer, 2009.

[4] K. W. 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*, 101(1):1–15, 2006.

[5] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama. 3D face recognition under expressions, occlusions, and pose variations. *PAMI*, 2013.

[6] J.-M. Fellous. Gender discrimination and prediction on the basis of facial metric information. *Vision research*, 37(14), 1997.

[7] S. Fu, H. He, and Z. Hou. Learning race from face: A survey. *PAMI*, 2014.

[8] G. Guo. Human age estimation and sex classification. In *Video Analytics for Business Intelligence*. Springer, 2012.

[9] Y. Hu, J. Yan, and P. Shi. A fusion-based method for 3D facial gender classification. In *ICCAE*, 2010.

[10] G. Mu, G. Guo, Y. Fu, and T. S. Huang. Human age estimation using bio-inspired features. In *CVPR*, 2009.

[11] T. Russ, C. Boehnen, and T. Peters. 3D face recognition using 3D alignment for pca. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1391–1398. IEEE, 2006.

[12] R. Tokola, D. Bolme, D. Barstow, C. Boehnen, and K. Ricanek. Discriminating projections for estimating face age in wild images. In *IJCB*. 2014.

[13] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1), 1991.

[14] B. Xia, B. B. Amor, M. Daoudi, H. Drira, et al. Can 3D shape of the face reveal your age? In *VISAPP*, 2014.

[15] C. Xu, Y. Wang, T. Tan, and L. Quan. A new attempt to face recognition using 3D eigenfaces. In *ACCV*. Citeseer, 2004.

[16] W. Yang, C. Chen, K. Ricanek, and C. Sun. Ensemble of global and local features for face age estimation. In *Advances in Neural Networks*. 2011.

[17] C. Zhang and Z. Zhang. A survey of recent advances in face detection. Technical report, Technical report, Microsoft Research, 2010.