International Journal of Computer Vision manuscript No.
(will be inserted by the editor)

Ethnicity- and Gender-based Subject Retrieval Using 3-D Face-Recognition Techniques

George Toderici[†] · **Sean M. O'Malley**[†] · **George Passalis**[†,‡] · **Theoharis Theoharis**[†,‡] · **Ioannis A. Kakadiaris**[†]

Received: date / Accepted: date

Abstract While the retrieval of datasets from human subjects based on demographic characteristics such as gender or race is an ability with wide-ranging application, it remains poorly-studied. In contrast, a large body of work exists in the field of biometrics which has a different goal: the recognition of human subjects. Due to this disparity of interest, existing methods for retrieval based on demographic attributes tend to lag behind the more well-studied algorithms designed purely for face matching. The question this raises is whether a face recognition system could be leveraged to solve these other problems and, if so, how effective it could be. In the current work, we explore the limits of such a system for gender and ethnicity identification given (1) a ground truth of demographically-labeled, textureless 3-D models of human faces and (2) a state-of-the-art face-recognition algorithm. Once trained, our system is capable of classifying the gender and ethnicity of any such model of interest. Experiments are conducted on 4007 facial meshes from the benchmark *Face Recognition Grand Challenge v2* dataset.

Keywords ethnicity, face, gender, identification, race, recognition, retrieval.

[†]Computational Biomedicine Lab
Department of Computer Science
University of Houston
4800 Calhoun
Houston, TX 77204-3010
E-mail: {gtoderici,somalley,ioannisk}@uh.edu
.
[‡]Computer Graphics Laboratory
Department of Informatics & Telecommunications
University of Athens
Panepistimiopolis
15784 Athens, Greece
E-mail: {passalis,theotheo}@di.uoa.gr

1 Introduction

Face recognition is an intensely-studied task in computer vision. There exist a plethora of algorithms which address this problem for 2-D intensity images, 2.5-D images (i.e., range scanners), 3-D meshes with and without textural information, and the fusion of these and any number of more exotic modalities. Face recognition research has also branched in a limited manner to address related problems such as the estimation of age, gender, and ethnicity based on features measured from the human face. Such abilities have broad application in identity verification, criminal forensics, anthropology, and other fields which rely on accurate anthropometry. Unfortunately, these non-recognition problems remain less popular than the extremely well-studied problem of identity verification.

In this paper, we address two of the above problems: the estimation of gender and race[1] based on facial imagery. We will be examining 3-D meshes of the face without any associated texture or photographic information. As skintone will be ignored, these experiments provide a demonstration of the discriminative power of facial structure alone. However, instead of explicitly extracting features from the facial model in an attempt to capture the gender or ethnicity of the face, we leverage prior research into face recognition to accomplish the same task. In this paper, we assume a face recognizer is some function which provides a distance $d(x, y) \geq 0$ between two subjects x and y, where ideally $d(x, x) \approx 0$ and $d(x, y) > d(x, x)$ for $x \neq y$. We seek to answer if such a function may act as a proxy for a higher-level application such as gender or race classification. Superficially this may not appear to be the case: among other

[1] We use "race" and "ethnicity" interchangeably here as the same concept is expressed by both terms in the prior literature.

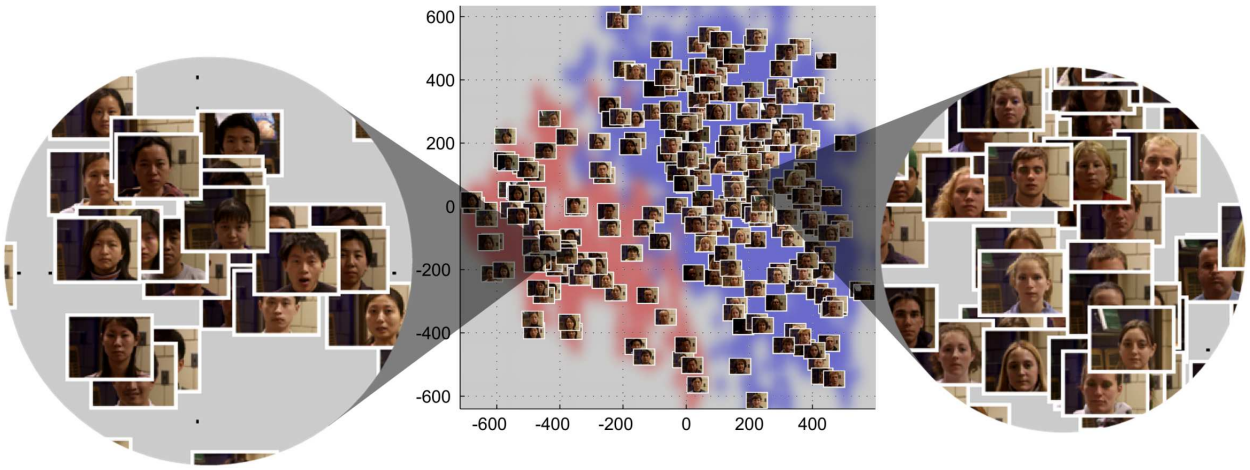


Fig. 1 A depiction of the similarity space to be described later: the central plot contains clouds of points organized automatically by our algorithm, labeled according to our ground truth (red=*Asian*, blue=*White*), and smoothed to highlight the segmentation of the classes. A random subset of our subjects' photographs have been drawn in their appropriate locations as an overlay and magnified in the two circular figures. This figure illustrates the natural clustering into racial groups which is a byproduct of our method. Projected along other dimensions, the separation of genders may be witnessed as well. While differences in skin tone are obvious here, note that our method used only facial shape to generate the data for this plot: no skin texture or other photographic data were involved. The meaning of the axes of this plot will be discussed in Sec. 3.3.3.

reasons, face recognition algorithms are notoriously sensitive to intra-subject facial variation.

In this paper, we start with a groundtruth dataset of n meshes, for each mesh i of which we are provided a gender g_i and a racial category r_i. For a face distance function we use one earlier developed by our group (Kakadiaris et al (2007)). Then, given this data and this function, the tasks before us are to determine the most likely gender and ethnicity of any given subject (i.e., facial mesh). One of the techniques we present to solve this problem makes partial use of an automatically-constructed space in which subjects similar in appearance occupy localized regions in the space (Fig. 1). Our main contributions are (1) an analysis of the effectiveness of a purely facial-structure-based distance function for gender and ethnicity classification, (2) a training scheme which is agnostic of the underlying facial distance function, and (3) a resulting fully-automatic system which achieves accuracy of $\sim 99\%$ for race and $\sim 94\%$ for gender on a public benchmark dataset. (For context, a comparison to what we believe to be the most similar existing system will be presented in the following section.)

2 Background

It has been argued for some time that the 3-D structure of the face is a more effective indicator of gender than facial texture (O'Toole et al (1995, 1997)). Recent research has developed this idea to provide gender detectors based on 3-D models of the face (Lu et al (2006)) and from 2.5-D models of the face derived from shape-from-shading algorithms (Wu et al (2007, 2008)). However, since extant facial imagery still predominantly consists of 2-D intensity images, researchers continue to develop algorithms for gender detection in these data (Gutta et al (2000); Lian et al (2005); Yang and Ai (2007)). For processing very large datasets, these algorithms have been tuned to operate on extremely low-resolution "thumbnail" images (i.e., 24 pixels or less along one axis) (Moghaddam and Yang (2002); Baluja and Rowley (2007); Mäkinen and Raisamo (2008)). A surprising commonality between these studies is that increasing image resolution leads to little improvement in performance when the imagery is normalized with respect to lighting and facial alignment. With some exceptions (e.g., Baluja and Rowley (2007)), these studies also show consistently good performance from support vector machine (SVM) learners.

The discrimination of ethnicity from facial imagery remains relatively undeveloped compared to the corresponding algorithms for gender identification. Some recent works have looked at this task in the context of intensity images (Hosoi et al (2004); Lu and Jain (2004); Yang and Ai (2007)) and 3-D range imagery (Lu et al (2006)). The latter work suggests, similarly to the earlier case of gender, that 3-D information can be by itself superior to intensity information for the identification of ethnicity. With the exception of Hosoi et al (2004), which considers *African*, *Asian*, and *European* classes, all of the cited works consider a binary classification problem: *Asian* versus *non-Asian*. This is not

necessarily due to algorithmic limitations, but to a lack of standard datasets which contain significant representation from other classes.

A concept similar to the "face-similarity space" used for illustrative purposes later has been described in the context of analyzing facial attractiveness (Potter et al (2007)). Such mappings have a long tradition in the visualization of human preference data (Young (1987)). However, the results obtained in such preference studies tend to require extensive manual effort (by definition), making such approaches impractical for database-scale use. Computational methods are therefore becoming more common for these tasks. These methods make use of multidimensional scaling (MDS) and related embedding algorithms not only for analyzing (down-projecting) 3-D facial mesh data onto simpler domains, but also for inter-face comparison (Elbaz and Kimmel (2003); Bronstein et al (2006, 2007)) and representing raw distances in a more easily-visualized Euclidean space (Aharon and Kimmel (2006); Elbaz and Kimmel (2003)). The latter is especially relevant to the procedure described later (Sec. 3.3.3).

In the current work, the only manually-produced data required are race and gender labels over a set of training meshes. During deployment, only the mesh of the target subject is required. The most similar system to this we are aware of is that of Wu et al (2007), though instead of operating on 3-D meshes, the authors employ 2.5-D needle maps recovered with shape-from-shading. That study reports a gender recognition level of 93.6% on 260 manually-aligned images (the study does not address race). This performance is comparable to our own system, however we do not require manual guidance.

3 Methods

3.1 Data

For this study we use the set of 3-D facial meshes made available by the *Face Recognition Grand Challenge v2* (Phillips et al (2005)). These meshes were captured with a commercial structured light sensor. For the purposes of these experiments we ignore texture information and associated still photographs (except for later illustration).

Metadata for the meshes in this dataset include the following racial categories: *Asian* ($n = 1121$), *Asian-Middle-Eastern* ($n = 16$), *Asian-Southern* ($n = 78$), *Black* ($n = 28$), *Hispanic* ($n = 113$), *Unknown* ($n = 97$), and *White* ($n = 2554$). Included gender labels are *Female* ($n = 1840$) and *Male* ($n = 2167$). Each subject participated in from 1 to 22 separate imaging sessions; the 4007 total meshes were captured from 466 subjects. For the race-determination experiments in this paper we will consider only the *Asian* and *White* classes, as the others contain too few participants to support meaningful results. However, the methods presented in this paper are not inherently binary and can readily be extended to multiple classes where the training data is sufficient to do so. For our gender experiments, all subjects are included regardless of their racial category.

As in the face-recognition literature, the ground-truth dataset will be referred to as the *gallery*. The unlabeled face mesh for which we are tasked with determining gender and ethnicity will be referred to as the *probe*.

3.2 Face Distance Measure

A detailed description of our 3-D face recognition system, URxD, is not necessary here (see Kakadiaris et al (2007)), but we will describe it briefly.[2] The main idea of our approach is to represent an individual's facial structure as a deformed version of a "standard" human face. The deformed model captures the idiosyncracies of the specific face and represents its 3-D geometry in an efficient 2-D structure by utilizing the model's UV parameterization. This structure is decomposed using both Haar wavelet decomposition and the steerable pyramid transform (Simoncelli et al (1992)). The two resulting sets of coefficients define the final metadata that are used for comparing different subjects.

Given the metadata for a pair of subjects, we may then define a distance function between the two geometries. This is obtained as a weighted sum of two independent distance measures: an L^1 measure on the Haar wavelets and the CW-SSIM similarity measure on the pyramid coefficients (the latter is a translation-insensitive similarity measure inspired by the structural similarity (SSIM) index of Wang et al (2004)).

It is important to note that for the purposes of the current work, the absolute distances between facial meshes are not as relevant as the idea that our distance function should correspond to the intuitive concept of a facial distance measure described earlier (Sec. 1). Namely, that for our measure $d(\cdot, \cdot)$, if $d(\mathbf{a}, \mathbf{b}) < d(\mathbf{a}, \mathbf{c})$ then meshes $\mathbf{a}$ and $\mathbf{b}$ are more likely to belong to the same subject than $\mathbf{a}$ and $\mathbf{c}$. Under this assumption, face recognition algorithms of sufficient strength may in principle be used interchangeably for the task at hand. How this is accomplished will be described in the following sections.

[2] This algorithm competed in the 2006 *Face Recognition Vendor Test* and achieved the top performance in the shape-only (3-D) category.

3.3 Gender/Ethnicity Estimation

Assuming we are given a ground-truth gallery of n facial meshes $\mathbf{x}_i$ (not necessarily from unique subjects) along with associated gender g_i and race r_i information, we are tasked with the problem of, given a probe mesh $\mathbf{x}_{n+1}$, to determine its most probable labels g_{n+1} and r_{n+1}. For the purpose of comparing the performance of successively more advanced techniques for accomplishing this, we will present four possible methods along with their performance tradeoffs. Experimental results for each will be presented in Sec. 4.

3.3.1 Solution #1: k-Nearest-Neighbors

The most obvious solution to our task is to find the k most similar meshes to our probe mesh and, through majority voting on their gender and ethnicity labels, pick g_{n+1} and r_{n+1}. (In cases where ethnicity is a non-binary decision, k must be extended incrementally until a clear winner emerges.)

3.3.2 Solution #2: Kernelized k-Nearest-Neighbors

One apparent shortcoming of the previous approach is it lacks consideration of the absolute distances in the nearest-neighbor list. By applying a weight function that decays by distance, we may remedy this. For example:

$$w_{\text{male}} = \sum_{\mathbf{x} \in \text{males}} \exp(-\sigma d(\mathbf{x}_{\text{probe}}, \mathbf{x})), \tag{1}$$

where $\mathbf{x}_{probe}$ is the probe mesh, σ is a falloff parameter, $d(\cdot,\cdot)$ our face-similarity function, and w_{male} can be considered a confidence score that the subject belongs to the *male* class. This function may be modified in the obvious manner to score competing classes. The optimal value of σ will vary according to the distance function used; for the results presented later we determined through a grid search that $\sigma = 1/8$ resulted in the highest classification accuracy.

3.3.3 Solution #3: Learning from the Face-Similarity Space

The previous naive techniques may be effective to a degree, but they make little attempt to understand the relationships between any faces in our gallery except for the probe. We now describe a more elaborate method which, during a training phase, constructs a *face similarity space* from our gallery. A high-level learning algorithm segments this Euclidean space into subregions which are intended to be occupied by only a single demographic label (i.e., based on the training set). This is performed twice: once for gender and once for ethnicity. During deployment, the location of the probe within the space is determined and its coordinates treated as a feature vector. With the aid of the previously-learned models, the demographic labelings corresponding to the location of this new point are determined. For the results we present in this paper, we use off-the-shelf support vector machines (SVMs) with their parameters optimized for our tasks of gender and ethnic identification.

Similarity space construction. In this step, we construct a face-similarity space from our gallery. In this space, each face will be represented as a point in a Euclidean space of p dimensions, where $p \ll n$, and the distance between each pair of points approximates the inter-face distances $d(\cdot,\cdot)$ which are the product of our face recognition algorithm. We begin by organizing the inter-face distances into the symmetric distance matrix

$$\mathbf{D} = \begin{pmatrix} d_{1,1} & d_{1,2} & \cdots & d_{1,n} \\ d_{2,1} & & & \vdots \\ \vdots & & \ddots & \\ d_{n,1} & \cdots & & d_{n,n} \end{pmatrix}, \tag{2}$$

where $d_{i,j} = d(\mathbf{x}_i, \mathbf{x}_j)$ and $d_{i,i} = 0$. We use multidimensional scaling (MDS) (Härdle and Simar (2003); Kruskal and Wish (1978); Seber (1984)) to transform this distance matrix into the desired point cloud. This is accomplished by letting $\mathbf{A}$ be the matrix where $a_{i,j} = -\frac{1}{2} d_{i,j}^2$ and letting $\mathbf{C}_n$ be the $n \times n$ centering matrix, $\mathbf{C}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$ (where $\mathbf{I}$ is the identity and $\mathbf{1}$ is a column vector of unit entries). We let $\mathbf{B} = \mathbf{C}_n\mathbf{A}\mathbf{C}_n$ and let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ and $\mathbf{v}_1 \ldots \mathbf{v}_n$ be the eigenvalues and associated eigenvectors of $\mathbf{B}$. The number of positive eigenvalues is denoted by $p \leq n$. We now form the matrix

$$\mathbf{Y} = \left(\sqrt{\lambda_1}\mathbf{v}_1, \sqrt{\lambda_2}\mathbf{v}_2, \ldots, \sqrt{\lambda_p}\mathbf{v}_p \right), \tag{3}$$

where each row of $\mathbf{Y}$ specifies a point in a p-dimensional space. The i^{th} row of $\mathbf{Y}$ corresponds to the i^{th} face in our gallery and collectively $\mathbf{Y}$ populates our face-similarity space.

Similarly to methods used in principal component analysis, we may reduce the dimensionality of the space described by $\mathbf{Y}$ to a number of dimensions $p' < p$ to make subsequent computations less expensive. The rightmost $p - p'$ columns of $\mathbf{Y}$ may then be removed. This may be accomplished in two ways. For the first, we define the function $f(i) = \sum_{j=1}^{i} \lambda_j \Big/ \sum_{j=1}^{p} \lambda_j$ and let p' be the minimum integer such that $f(p') \geq \beta$, where $1 \leq p' \leq p$ and $\beta \in [0,1]$ is a retention threshold (i.e., values closer to 1 lead to higher p'). However, this only

considers the general importance of each axis without considering the actual points in the space. A more informative method is to analyze the *stress* of the point configurations induced by various p'. Stress is measured by producing a dissimilarity matrix $\mathbf{D}^Y$ from $\mathbf{Y}$ itself (i.e., by measuring the Euclidean distance between each pair of rows in $\mathbf{Y}$) and comparing $\mathbf{D}^Y$ to the original matrix $\mathbf{D}$ by a measure such as

$$S(\mathbf{D}, \mathbf{D}^Y) = \sqrt{\frac{\sum_i \sum_j (d_{i,j} - d^Y_{i,j})^2}{\sum_i \sum_j d^2_{i,j}}}. \quad (4)$$

As fewer of the rightmost columns of $\mathbf{Y}$ are retained, we expect S to increase. Selecting a p' associated with an acceptable level of stress again allows the dimensionality of our space to be reduced. Note, though, that the MDS transformation can only be zero-error if the distances in $\mathbf{D}$ are Euclidean (equivalently, if $\mathbf{B}$ is positive semidefinite). In our case this is not true, so stress will be non-zero. In practice, stress and the selection of p' is somewhat dependent on the idiosyncracies of the face-recognition algorithm providing the distance function that is the basis of $\mathbf{D}$. However, as we will see later, p' may be small compared to n while still providing reasonable results.

For the remainder of this paper, we will assume $\mathbf{Y}$ is an $n \times p'$ matrix, where n is the number of faces in our gallery and $p' \ll n$ is the p'-dimensional location of each face. These p'-vectors may be considered a compact representation of the original facial meshes, though, of course, these feature vectors are only relevant in the context of the entire gallery.

Determining the classification of a probe face. Given an unclassified probe mesh $\mathbf{x}$, we first determine the p'-vector $\mathbf{v}$ which places $\mathbf{x}$ in our space $\mathbf{Y}$. To begin, we find the distances between $\mathbf{x}$ and m randomly-chosen faces in our gallery, $\mathbf{f}_{1..m}$. The higher m, the greater the accuracy of our placement, but generally this value can be much smaller than the size of the gallery.

Next, the initial location of $\mathbf{v}$, $\mathbf{v}_0$, is set to the p'-length zero-vector: $\mathbf{v}_0 = \mathbf{0}^{p'}$ (incidentally, this places $\mathbf{v}_0$ at the center of mass of our space $\mathbf{Y}$). Lastly, a simple gradient descent moves $\mathbf{v}_0$ into its final location. The cost function we minimize during this process is equivalent to the stress function (4) which was globally minimized during the construction of $\mathbf{Y}$:

$$\mathbf{v} = \min_{\mathbf{v}_0} \sqrt{\sum_{i=1}^{m} [d(\mathbf{x}, \mathbf{f}_i) - \|\mathbf{v}_0 - \mathbf{y}_i\|]^2}, \quad (5)$$

where $\mathbf{v}$ is the final estimate of the probe face's location given its starting point $\mathbf{v}_0$, $\mathbf{x}$ is the probe mesh, $\mathbf{f}_{1..m}$ are the randomly-selected gallery faces, and $\mathbf{y}_{1..m}$ are their locations in the similarity space (corresponding to rows in $\mathbf{Y}$). The vector $\mathbf{v}$, along with the race and gender models learned previously, together provide the final classification of our probe face.

Computational complexity. The complexity of the operation described by (2) is $n(n-1)/2 = \mathrm{O}(n^2)$, while our earlier nearest-neighbor solutions were approximately $\mathrm{O}(n)$. There are two ways to minimize this computational cost: using a simpler face distance measure $d(\cdot, \cdot)$, and sparsifying $\mathbf{D}$. We will ignore the first option as this puts severe limitations on which face recognition algorithms may be used for our task. As for making $\mathbf{D}$ sparse, unlike spectral clustering approaches such as normalized cuts (Shi and Malik (2000)), classical MDS does not allow us to ignore entries in our distance matrix. This is unfortunate as there is a tremendous amount of redundancy in such matrices: it is not difficult to find reasonable values for missing distances based on the remaining associations of each point to its neighbors. However, nonmetric multidimensional scaling approaches exist which allow the creation of a similarity space from incomplete information (Tsogo et al (2000)). It is not necessary to labor this point here; suffice it to say that if $d(\cdot, \cdot)$ is sufficiently complex, much of the burden of filling $\mathbf{D}$ can be eliminated in lieu of moderately greater construction cost and error in $\mathbf{Y}$. The use of a sparse $\mathbf{D}$ ultimately obviates the need for a significant number of inter-face comparisons; in fact, $\mathbf{D}$ may potentially be constructed as a band-diagonal matrix.

3.3.4 Solution #4: Learning from Algorithm-Specific Features

One of the premises of this work is that any sufficiently advanced face-similarity measure may be employed as an interchangeable element in a system for identifying and/or retrieving face imagery based on its demographic characteristics. One way to accomplish this was described previously (Sec. 3.3.3). A point of interest this raises, though, is how much of a decrease in performance do we suffer when we insist that the face-similarity algorithm itself be treated as a black box. To answer this question, we now relax this constraint.

As described earlier (Sec. 3.2), one of the byproducts of our face recognition algorithm is a set of wavelet coefficients which compactly describe the shape of the face. These coefficients are derived from the geometry-image representation of the fitted deformable model. This description of the face is far richer than our previous, which represents each face as a point in a relatively low-dimensional Euclidean space. As such, we

would expect to obtain higher performance using these coefficients. To accomplish this, we revise our previous solution by exchanging the wavelet coefficients for the p'-dimensional location of each point. All other steps of the algorithm remain the same. That is, instead of our training data consisting of n pairs $\langle \mathbf{x}_{1..n}, class_{1..n} \rangle$ where $\mathbf{x}_i$ is a face's location in a face-similarity space and $class_i$ its associated class (e.g., *Male* or *Female*), we substitute $\mathbf{x}_i$ for the face's wavelet coefficients and proceed with training and deployment as usual.

4 Results

4.1 Combined Gender/Race Retrieval

For a retrieval scenario in which the user is interested in retrieving meshes based on both gender and race, we combine the probability estimates from the race and gender SVMs. These two SVMs use 3^{rd}-degree polynomial kernels and are optimized through independent grid searches. Our underlying SVM implementation is libSVM (Chang and Lin (2001)).

As the outputs of both classifiers are probabilistic, these outputs may be multiplied to obtain a joint probability. The user may then choose an operating point (i.e., a probabilistic threshold) in order to retrieve all meshes matching the desired criteria. The resulting ROC curves for each possible race/gender retrieval combination are illustrated in Fig. 2(a) for the MDS technique (Sec. 3.3.3) and Fig. 2(b) for the wavelet (Sec. 3.3.4). Here, cross-validation (10-fold) was employed while ensuring that the meshes for a specific subject are used in either training or testing, but not both. The relevant performance metrics from all folds were combined, thus providing our summary ROC curves.

Note that this particular experiment excludes subjects belonging to racial categories which lack adequate representation in our gallery, so only 3676 of the 4007 available meshes were used to generate these curves. (The gender-specific results presented in the following sections include the complete gallery.)

4.2 Independent Gender/Race Classification

We present the results of four different types of classification experiments: (1) k-nearest-neighbors (kNN) (here, a majority vote among the $k = 10$ most similar faces, Sec. 3.3.1), (2) kernelized kNN (k-kNN, Sec. 3.3.2), (3) learning based on the face-similarity space (MDS, Sec. 3.3.3), and (4) learning based on wavelet coefficients (Sec. 3.3.4). In the last two cases, 10-fold cross-validation is performed for training and testing

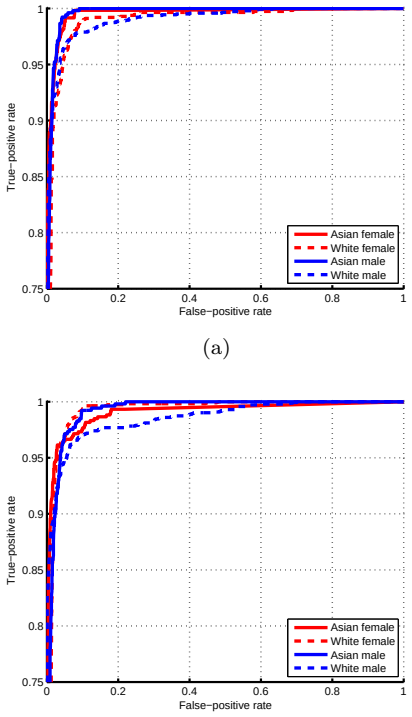


Fig. 2 ROC curves for the combined gender/race retrieval task using (a) the MDS representation of our gallery meshes and (b) the wavelet-coefficient representation. (Note that the y-axis has been truncated below 0.75 for clarity.)

and the mean±standard deviation in accuracy across these 10 runs is reported in the confusion matrices to be presented later. (The first two cases do not explicitly use machine learning, so the cross-validation approach is not necessary in that case.) That is, $9/10^{\text{th}}$ of the labeled meshes are used for training each fold and the meshes in the remaining $1/10^{\text{th}}$ of the data are used as probes to test the trained models.

We limit our MDS space to 150 dimensions; equivalently, each face is described by a 150-dimensional vector for experiments of type (3). The number of wavelet coefficients used for experiment (4) is 3608. As in our earlier experiments, SVM learners are used for experiments (3) and (4), which differ only in the type of feature vector used to describe each face. The demographic labels associated with each face are not used during construction of the similarity space for experiment (3).

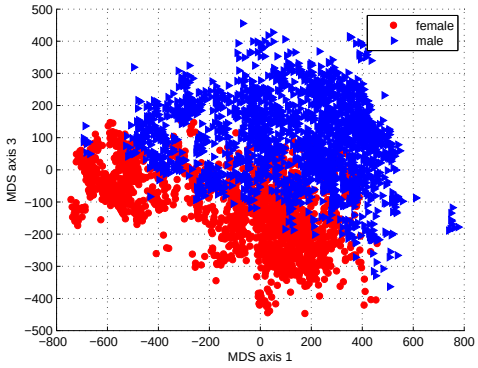
We present our results for gender identification in Table 1 and for race in Table 2.

		Male	*Female*
kNN	*Male*	92%	18%
	Female	8%	82%
k-kNN	*Male*	93%	18%
	Female	7%	82%
MDS	*Male*	93.3% ± 6%	7.7% ± 5%
	Female	6.7% ± 6%	92.3% ± 5%
Wavelets	*Male*	94% ± 5%	7% ± 4%
	Female	6% ± 5%	93% ± 4%

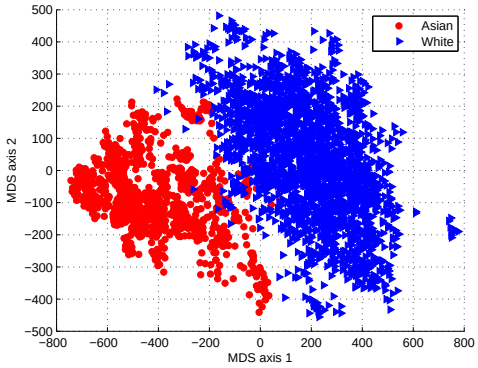
Table 1 Confusion matrices for the four methods (kNN, kernelized kNN, face-similarity space, and wavelet features) for determining gender. Values are the mean accuracy over all subjects. The left column indicates the ground-truth labels, the top row the predicted labels. For those experiments which require cross-validation, $x \pm y$ indicates the mean and standard deviation in accuracy over 10 folds.

		White	*Asian*
kNN	*White*	99.1%	1.6%
	Asian	0.9%	98.4%
k-kNN	*White*	99.1%	1.6%
	Asian	0.9%	98.4%
MDS	*White*	99.6% ± 0.01%	0.5% ± 0.1%
	Asian	0.4% ± 0.1%	99.5% ± 0.1%
Wavelets	*White*	98.2% ± 2%	2.9% ± 3%
	Asian	1.8% ± 2%	97.1% ± 3%

Table 2 Confusion matrices of the type presented in Table 1, but expressing racial classification. Classes which are poorly represented in our data are excluded (see Sec. 3.1).



(a)



(b)

Fig. 3 Face-similarity space projected into two dimensions and labeled according to (a) gender and (b) ethnicity. The projection axes were chosen for maximum visual class separation.

4.3 Other Results

A byproduct of our algorithm is a face-similarity space which, by visual inspection, illustrates interesting features of our dataset. In Fig. 3, for instance, we show our space labeled according to race and gender. As we may observe, even though the space is built without regard for these demographic labels, it is not difficult to visually separate the classes even when most of the space's dimensions have been eliminated: here we illustrate only 2 of the 150 dimensions used previously for classification.

In Figs. 4 and 5, we collect photographs from subjects in our dataset by projecting the similarity space along each of two axes and sampling faces at intervals. We can see, for instance in Fig. 4, the progression of faces from "definitely female" to "definitely male." (Of course, since this is a very rudimentary way of separating these classes, not all female subjects appear prior to male subjects.)

As the centroid of a class's point cloud corresponds to the location that is, on average, the lowest distance from all other points, it may be of interest examining not only the faces closest to these locations, but also those furthest away. In Fig. 6 we illustrate this for the gender identification problem and in Fig. 7, render the corresponding meshes.

5 Conclusion

We have discussed a number of approaches for leveraging existing face-recognition technologies for the task of subject retrieval based on high-level demographic features (gender and race) estimated from 3-D meshes of the human face. Both our MDS and wavelet approaches provide high levels of classification performance on the benchmark dataset: $> 99\%$ mean accuracy for MDS on the race task and $\approx$ 94% for wavelets on the gender task. What renders these results especially unusual is the fact that our MDS approach is trained on feature vectors which are generated entirely without regard for demographic labels or even explicit knowledge of the facial structure of each subject. In addition, the technique is even agnostic of the underlying function which provides it with face-similarity distances. In spite of this, even off-the-shelf learning algorithms (in our case,

(a) -446.89 (b) -189.93 (c) -155.95 (d) -125.25

(e) -25.079 (f) -9.1402 (g) 9.819 (h) 26.527

(i) 161.25 (j) 191.29 (k) 228.51 (l) 273.25

Fig. 4 Photographs of subjects sampled along the dimension most discriminative of gender in our data (dimension 3 in Fig. 3(a)).

(a) -743.12 (b) -580.63 (c) -522.05 (d) -451.01

(e) 28.413 (f) 66.656 (g) 101.54 (h) 137.46

(i) 316.63 (j) 363.72 (k) 396.51 (l) 453.62

Fig. 5 Photographs of subjects sampled along the dimension most discriminative of race in our data (dimension 1 in Fig. 3(b)).

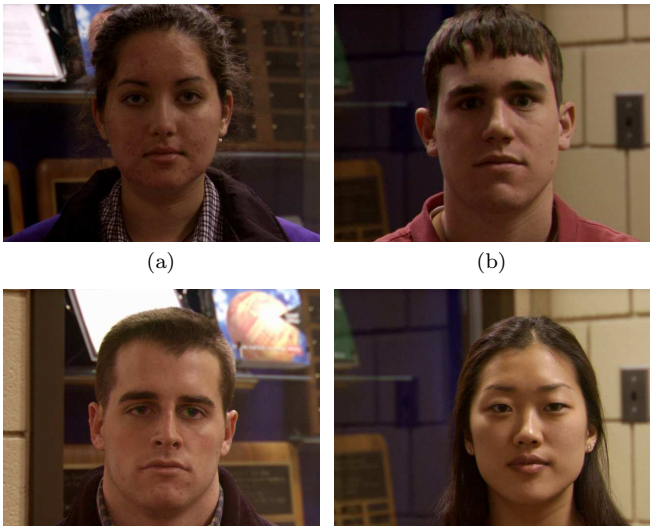


Fig. 6 The "most typical" faces in the gallery: (a) the female face closest to the female centroid in the similarity space and (b) the male face closest to the male centroid. Outlier faces: (c) the male face furthest from the female centroid and (d) the female face furthest from the male centroid.

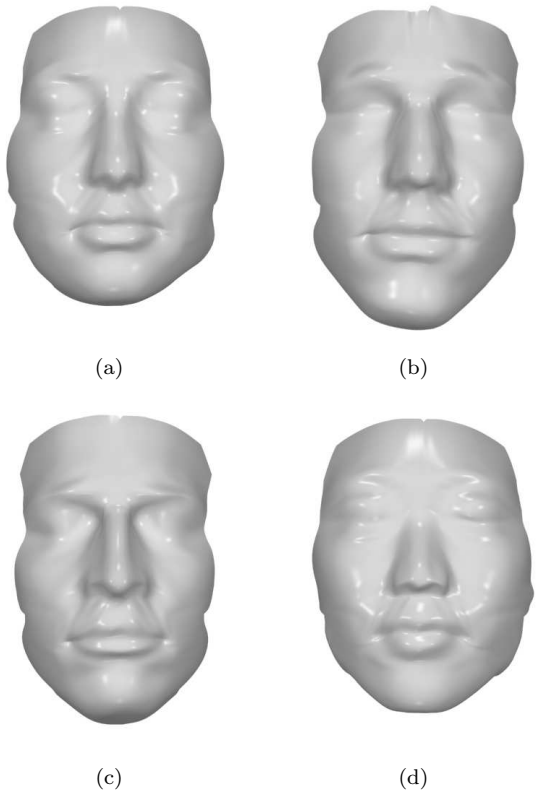


Fig. 7 Rendered meshes corresponding to the faces in Fig. 6. These are not the original laser-scanned images, but the deformable meshes after fitting to the range data as part of our recognition algorithm (Sec. 3.2). (Note that model rotation and aspect ratio will not necessarily match the photographs.)

SVMs) trained on the face-similarity space are capable of surprisingly high levels of performance.

Interestingly, the proposed method of learning from the wavelet representation of the face, which we expected to outperform MDS, actually performs worse on the race-classification task. There are three factors which could lead to this situation. For one, the "race" task is inherently fuzzier than the "gender" task, though both labels are treated as binary. (The race labels are self-reported by the participants as the race they most identify with.) Secondly, the wavelet representation is higher-dimensional than MDS. Lastly, it may suffer from greater noise as it does not benefit from the implicit noise reduction of MDS' dimensionality reduction. One way to alleviate these problems would be to increase the size of our training corpus; however, we are constrained by the bounds of the existing benchmark dataset.

6 Future Work

A natural alternative to our system of converting the gallery to a distance matrix, the distance matrix to the face-similarity space with MDS, and then learning from the point-cloud representation of the faces, is to use spectral clustering methods on the distance matrix itself. We have so far ignored this topic for two reasons: (1) by projecting into a Euclidean space, MDS is spectacularly well-suited to visualization, an ability spectral clustering does not share; and (2) spectral clustering's strength is in connected-component analysis, which is not necessarily the best choice for our data. However, we recognize that spectral techniques could be a fruitful ground for future discoveries in this area.

One limitation of the current work is a lack of data for ethnicities outside of *Asian* and *White*. As such, our experiments can only serve to illustrate the potential power of our approach for solving n-class retrieval problems. We hope this dearth of labeled facial data will be addressed by future acquisition studies.

While we have argued that demographic retrieval tasks can benefit from the vast existing body of work in face recognition, our results suggest possible areas of future research which would be mutually beneficial to both areas. For instance, we have observed that commonalities in human facial morphology due to race and gender express themselves in surprisingly compact subspaces in the universe of faces. As such, one possible research direction is the possibility of exploiting ensembles of race- or gender-specific face recognition machines, under the assumption that algorithms trained on individual subspaces would be better-tuned to their idiosyncracies than the current standard of training one

system to distinguish all faces. This concept we must leave as a target of future work.

Acknowledgements We would like to thank Michael Fang (U. of Houston) for his invaluable assistance in rendering the mesh figures for this paper.

References

Aharon M, Kimmel R (2006) Representation analysis and synthesis of lip images using dimensionality reduction. Int J Comp Vis 67(3):297–312

Baluja S, Rowley HA (2007) Boosting sex identification performance. Int J Comp Vis 71(1):111–119

Bronstein AM, Bronstein MM, Kimmel R (2006) Efficient computation of isometry-invariant distances between surfaces. SIAM J Scientific Computing 28(5):1812–1836

Bronstein AM, Bronstein MM, Kimmel R (2007) Expression-invariant representations of faces. IEEE T Image Process 16(1):188–197

Chang CC, Lin CJ (2001) LIBSVM: A library for support vector machines. Software available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm

Elbaz AE, Kimmel R (2003) On bending invariant signatures for surfaces. IEEE T Pattern Anal Mach Intell 25(10):1285–1295

Gutta S, Huang JJ, Jonathan P, Wechsler H (2000) Mixture of experts for classification of gender, ethnic origin, and pose of human faces. IEEE T Neural Networks 11(4):948–960

Härdle W, Simar L (2003) Applied Multivariate Statistical Analysis, 1st edn. Springer

Hosoi S, Takikawa E, Kawade M (2004) Ethnicity estimation with facial images. In: IEEE Int Conf Automatic Face and Gesture Recognition

Kakadiaris IA, Passalis G, Toderici G, Murtuza MN, Lu Y, Karampatziakis N, Theoharis T (2007) Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. IEEE T Pattern Anal Mach Intell 29(4):640–649

Kruskal JB, Wish M (1978) Multidimensional Scaling. SAGE Publications

Lian HC, Lu BL, Takikawa E, Hosoi S (2005) Gender recognition using a min-max modular support vector machine. In: Int Conf Advances in Natural Computation, pp 438–441

Lu X, Jain AK (2004) Ethnicity identification from face images. In: SPIE Int Symp Defense and Security, pp 114–123

Lu X, Chen H, Jain AK (2006) Multimodal facial gender and ethnicity identification. In: Int Conf Biometrics, Hong Kong

Mäkinen E, Raisamo R (2008) Evaluation of gender classification methods with automatically detected and aligned faces. IEEE T Pattern Anal Mach Intell 30(3):541–547

Moghaddam B, Yang M (2002) Learning gender with support faces. IEEE T Pattern Anal Mach Intell 24(5):707–711

O'Toole AJ, Vetter T, Bülthoff HH, Troje NF (1995) The role of shape and texture information in sex classification. Tech. rep., Max Planck Institut für biologische Kybernetik

O'Toole AJ, Vetter T, Troje NF, Bülthoff HH (1997) Sex classification is better with three-dimensional structure than with image intensity information. Perception 26:75–84

Phillips PJ, Flynn PJ, Scruggs T, Bowyer KW, Chang J, Hoffman K, Marques J, Min J, Worek W (2005) Overview of the Face Recognition Grand Challenge. In: IEEE Conf Comp Vis Patt Recog

Potter T, Corneille O, Ruys KI, Rhodes G (2007) "Just another pretty face": A multidimensional scaling approach to face attractiveness and variability. Psychonomic Bulletin & Review 14(2):368–372

Seber G (1984) Multivariate Observations, Wiley, chap 5.5: Multidimensional scaling

Shi J, Malik J (2000) Normalized cuts and image segmentation. IEEE T Pattern Anal Mach Intell 22(8):888–905

Simoncelli E, Freeman W, Adelson E, Heeger D (1992) Shiftable multi-scale transforms. IEEE T Inf Theory 38:587–607

Tsogo L, Masson MH, Bardot A (2000) Multidimensional scaling methods for many-object sets: A review. Multivar Behav Research 35(3):307–319

Wang Z, Bovik A, Sheikh H, Simoncelli E (2004) Image quality assessment: From error visibility to structural similarity. IEEE T Image Proc 13(4):600–612

Wu J, Smith WAP, Hancock ER (2007) Gender classification using shape from shading. In: British Machine Vision Conference

Wu J, Smith WAP, Hancock ER (2008) Facial gender classification using shape from shading and weighted principal geodesic analysis. In: Int Conf Image Anal Recog

Yang Z, Ai H (2007) Demographic classification with local binary patterns. In: Int Conf Biometrics, Seoul, Korea, pp 464–473

Young FW (1987) Multidimensional Scaling: History, Theory, and Applications. Lawrence Erlbaum Assoc