Posture Invariant Gender Classification for 3D Human Models

Stefanie Wuhrer   Chang Shu   Marc Rioux

National Research Council of Canada

Ottawa, Canada

Abstract

We study the behaviorally important task of gender classification based on the human body shape. We propose a new technique to classify by gender human bodies represented by possibly incomplete triangular meshes obtained using laser range scanners. The classification algorithm is invariant of the posture of the human body. Geodesic distances on the mesh are used for classification. Our results indicate that the geodesic distances between the chest and the wrists and the geodesic distances between the lower back and the face are the most important ones for gender classification. The classification is shown to perform well for different postures of the human subjects. We model the geodesic distance distributions as Gaussian distributions and compute the quality of the classification for three standard methods in pattern recognition: linear discriminant functions, Bayesian discriminant functions, and support vector machines. All of the experiments yield high classification accuracy. For instance, when support vector machines are used, the classification accuracy is at least 93% for all of our experiments. This shows that geodesic distances are suitable to discriminate humans by gender.

1. Introduction

Automatically determining the gender of a human is an important visual task that human beings can perform effortlessly. While gender classification based on two- and three-dimensional representations of human faces received considerable attention [17, 12, 16], little work considered gender classification based on the full 3D human body shape. The reason is that capturing and analyzing the full body shape is computationally costly.

With recent advances in laser range scanning technology, it becomes feasible to capture the full body shape. The Civilian American and European Surface Anthropometry Resource (CAESAR) data base [6] is a recent anthropometric survey that scanned the full body of about 4,500 civilians in North America and Europe using a 3D laser-range scanner. One product of the survey is the 3D location of 73
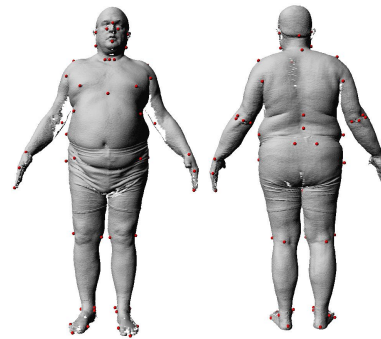

Figure 1. *Landmarks in the CAESAR data base.*

anthropometric landmarks for each of the scanned subjects. The landmarks were manually placed on the subjects prior to scanning and detected and identified in a semi-automated post-processing step. The 73 landmarks on one of the subjects are shown in Figure 1. The data of this survey allows us to analyze the human body shape and the influence of the body shape on automatic gender classification.

Godil et al. [11] use the CAESAR data base to demonstrate an algorithm for human identification based on few fixed distances. Lee and Grimson [15] examine gender classification of humans based on a representation called *gait*, which captures a human body in walking motion. Previous work that analyzed gender based on the full body shape concluded that the male and female body shapes differ [7]. However, gender classification was not the focus of this work.

We explore the problem of classifying humans by gender using the full human body shape represented by a triangular mesh. In particular, we are interested in finding the geodesic distances on the human body that are most suitable for gender discrimination. We examine the performance of three standard classification algorithms to classify by gender objects of the CAESAR data base. The features considered in this paper are based on all pairwise geodesic distances between anthropometric landmarks. Since geodesic distances on the human body surface are approximately posture invariant, our approach offers the advantage of classifying a human body shape independent of its posture. To our

knowledge, there is no prior work that classifies human beings according to gender that is invariant of the posture.

2. Data Used for Classification

We use data derived from the CAESAR data base to classify human body shapes by gender. Each subject of the CAESAR data base was scanned in three different postures. In this paper, we only consider subjects in the two postures shown in Figure 3. The reason is that the landmarks of the third posture are not yet available as the extraction process is laborious [6]. The body shapes are represented as triangular meshes. Each mesh is incomplete and contains between 200000 and 300000 vertices. Each body representation is enhanced by 73 landmarks.

The classifiers are trained using a *training set* of 500 female and 500 male subjects. The classifiers are then tested using a *test set* of 500 female and 500 male subjects. Clearly, the training set and the test set are disjoint. This paper only uses data acquired in North America.

Using the full mesh consisting of 200000 to 300000 vertices for classification is costly. Furthermore, using all of the mesh vertices for classification would require a one-to-one correspondence between all of the meshes in the data base. This correspondence is not readily available. We therefore only use all pairwise geodesic distances between landmarks for classification. Choosing geodesic distances as features offers the advantage of being approximately posture invariant. The reason is that geodesic distances on a human body are approximately preserved when the posture is changed as long as no topological changes occur.

Computing geodesic distances on a triangular mesh with holes is a challenging problem. Methods that walk through the triangles such as the fast marching method [14] and the method by Surazhsky et al. [19] yield inaccurate distances in this case. We therefore use the approach by Ben Azouz et al. [4] to find approximate geodesic distances between the landmarks. This approach maps the intrinsic geometry of the triangular mesh into $\mathbb{R}^d$ via multi-dimensional scaling and computes the geodesic approximations as Euclidean distances in $\mathbb{R}^d$. We choose $d = 3$ to compute the geodesic distance approximations.

Once all pairwise geodesics are computed, each subject of the CAESAR data base is represented as a point in $\mathbb{R}^{\frac{73 \cdot 72}{2}} = \mathbb{R}^{2628}$. Due to the curse of dimensionality, this feature size is too high. The following section outlines a step taken to reduce the feature size.

3. Dimensionality Reduction

Principal component analysis (PCA) is a common way to reduce the dimensionality of a data set. While PCA preserves the linear subspace with highest variance, it is not true that PCA is suitable to find the best subspace for discrimination [10].

We find a suitable subspace for discrimination of the feature space $\mathbb{R}^{2628}$ as follows. We compute the means M_m and M_f and covariance matrices $\sum_m$ and $\sum_f$ of the male and female training sets separately. The computation uses maximum likelihood estimates and is outlined below in Section 4.1. In the following, we only consider feature i if

$$|M_{m,i} - M_{f,i}| - 2\max(\sigma_{m,i}, \sigma_{f,i}) > 0, \qquad (1)$$

where $M_{m,i}$ and $M_{f,i}$ denote the i-th components of the means, and $\sigma_{m,i}$ and $\sigma_{f,i}$ denote the standard deviations along the i-th principal axis. The reason is that two distributions with $\sigma_{m,i} = \sigma_{f,i} = \sigma$ and $|M_{m,i} - M_{f,i}| - 2\sigma \leq 0$ are unimodal as shown by Helguero [9] and translated by Schilling et al. [18]. Those components are therefore not suitable for classification. The training data set reveals that only 15 features are not unimodal and potentially useful for classification.

We sort the features by the value of $|M_{m,i} - M_{f,i}| - 2\max(\sigma_{m,i}, \sigma_{f,i})$ in decreasing order and choose the first d features, such that $|M_{m,i} - M_{f,i}| - 2\max(\sigma_{m,i}, \sigma_{f,i}) \geq 0.001$. The ordering of the features is by importance for classification, since features where the distance between the means is large compared to the standard deviations of the distributions are important for classification.

Once this suitable subspace is chosen for classification based on the training data, both the training and test data sets are transformed into the new d-dimensional subspace. In our experiments d equals 11. That is, only a subgraph consisting of 11 of the 2628 original geodesic distances is used for classification. Figure 2 shows the graph on one of the CAESAR training data sets. Note that the geodesic paths are computed using the Dijkstra algorithm for the figure, since the approach used to approximate the geodesics does not compute a path [4]. All of the 11 components relevant for classification are located on the upper body. More precisely, the components describe the geodesic distances between the lower back and the face and the geodesic distances between the chest and the wrists. Note that for all of the geodesic distances selected for classification, the mean of the distances for male subjects is larger than the mean of the distances for female subjects.

4. Classification

This Section presents three methods to train classifiers in order to classify a given unknown sample. We first discuss how we estimate parameters using some assumptions about the data. We then apply three standard classification techniques to the data. The first classification finds the best linear classifier between the two classes. The second classification technique finds the Bayesian decision boundary
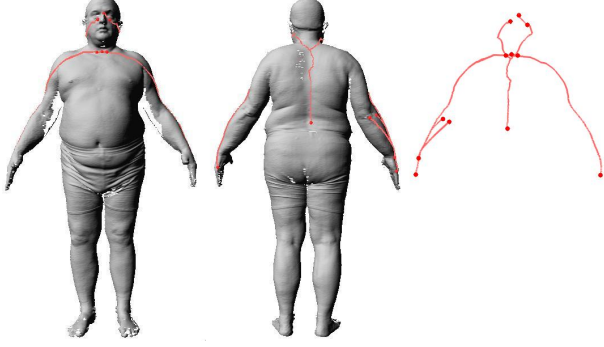
Figure 2. *Subgraph used for classification shown on one of the CAESAR models.*

between the two learned classes. The third classifier transforms the data into a high-dimensional space using a non-linear transformation and finds the best linear classifier in this space.

4.1. Parameter Estimation

We assume that the type of distribution of the data is known to be Gaussian. This assumption is inspired by anthropometric surveys [1, 2, 3] as well as previous research on the CAESAR landmarks [5]. For the training data sets, the prior probabilities for male and female measurements are $\frac{1}{2}$ each, since we extracted the d measurements for 500 male and 500 female training subjects from the CAESAR data base.

More precisely, we assume that the data points $\vec{x}_m$ and $\vec{x}_f$ have the following distributions:

$$d(\vec{x}_m) = \frac{1}{\sqrt{(2\pi)^d |\sum_m|}} e^{-\frac{1}{2}(\vec{x}_m - M_m)^T \sum_m^{-1} (\vec{x}_m - M_m)}$$
(2)

$$d(\vec{x}_f) = \frac{1}{\sqrt{(2\pi)^d |\sum_f|}} e^{-\frac{1}{2}(\vec{x}_f - M_f)^T \sum_f^{-1} (\vec{x}_f - M_f)}$$
(3)

where M_m, M_f are the means and $\sum_m, \sum_f$ are the covariance matrices of the male and female distributions, respectively.

We use the maximum likelihood parameter estimation. The assumptions stated above are sufficient to compute the mean and covariance matrix of the male and female data based on the given measurements. We compute the estimated means $\hat{M}_i^{MLE}$ and covariance matrices $\hat{\sum}_i^{MLE}$ for $i \in \{m, f\}$ as follows [10]:

$$\hat{M}_i^{MLE} = \frac{1}{500} \sum_{k=1}^{500} \vec{x}_i$$
(4)

$$\hat{\sum}_i^{MLE} = \frac{1}{500} \sum_{k=1}^{500} (\vec{x}_i - \hat{M}_i^{MLE})(\vec{x}_i - \hat{M}_i^{MLE})^T$$
(5)

The estimated parameters can be used to diagonalize the data and classify unknown data points.

4.2. Linear Discriminant Functions

This section is adapted from Duda et al. [10] and outlines how to examine whether the data is linearly separable. When using linear classification methods, we assume that the form of the discriminant function is known to be linear, which means that the discriminant function is a hyperplane in d-dimensional space. We can therefore express the classification boundary as

$$f(\vec{x}) = \begin{bmatrix} 1 & \vec{x}^T \end{bmatrix} \vec{a} = 0,$$
(6)

where $\vec{x}$ is a vector in d-dimensional space and $\vec{a}$ is a vector in $(d + 1)$-dimensional space. Since we aim to classify the sets of male and female measurements, we can find a linear discriminant function by solving the linear program

$$f(\vec{x}_m) > 0$$

$$-f(\vec{x}_f) > 0$$

if the data can be separated by a hyperplane. Note that the linear program consists of 1000 constraints, since we have 500 training samples of each class. We implemented the Perceptron Rule to test if the training data is linearly separable. Unfortunately, this is not the case. We therefore searched for the best existing linear classifier. To solve this problem, the Ho-Kashyap algorithm was used. Instead of solving a linear program, the Ho-Kashyap algorithm aims to solve the overdetermined system of 1000 linear equations

$$f(\vec{x}_m) = b_m$$

$$-f(\vec{x}_f) = b_f$$

where b_m and b_f are constants by minimizing the least squares error. To solve the problem, the Moore-Penrose inverse of the matrix

$$\begin{bmatrix} 1 & \vec{x}_m^T \\ -1 & -\vec{x}_f^T \end{bmatrix}$$

is computed. If the constants b_m and b_f are chosen arbitrarily, the resulting linear discriminant function may not be a good classifier. We chose $b_i = \frac{n}{n_i}, i \in \{m, f\}$, where $n = 1000$ is the total number of samples and $n_m = n_f = 500$ is the number of male and female training samples, respectively. This choice yields a classifier that apart from a multiplicative constant is identical to Fisher's discriminant.

Alternatively, the linear discriminant can be found by maximizing the dual Lagrangian

$$Q(\vec{\alpha}) = \sum_{i=1}^{n} \vec{\alpha}_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \vec{\alpha}_i \vec{\alpha}_j y_i y_j \langle \vec{x}_i, \vec{x}_j \rangle$$

with respect to Lagrange multiplier $\vec{\alpha}$ subject to the constraints $\vec{\alpha}_i \geq 0$ and $\sum_{i=1}^{n} y_i \vec{\alpha}_i = 0$. Here, y_i equals 1 if $\vec{x}_i$ is a male training sample and y_i equals -1 otherwise. The expression $\langle \vec{x}_i, \vec{x}_j \rangle$ denotes the dot product between $\vec{x}_i$ and $\vec{x}_j$. The solution to this problem is the linear discriminant given in Equation (6). This function can also be expressed as a linear combination of the given training samples

$$f(\vec{x}) = \sum_{i \in S} y_i \vec{\alpha}_i \langle \vec{x}, \vec{x}_i \rangle + b, \qquad (7)$$

where S is the set of support vector indices and where b is the axis intercept of the hyperplane. This is called the *dual form* of the linear discriminant function. Usually, only a small number of $\vec{\alpha}_i$ are non-zero. The training samples corresponding to the non-zero $\vec{\alpha}_i$ are called *support vectors*. We mention this approach because of its relation to support vector machines.

4.3. Bayesian Decision Boundary

This section is adapted from Duda et al. [10] and outlines how to examine if the data can be classified using the Bayesian decision boundary for Gaussian distributions. As in Section 4.1, we assume that the distributions of $\vec{x}_i, i \in \{m, f\}$ are given by Equations (2) and (3). We can then find a quadratic discriminant function as the optimal Bayes decision boundary. Since the prior probabilities are $\frac{1}{2}$ each in our case, the formula for the Bayes decision boundary becomes

$$\vec{x}^T (\textstyle\sum_f^{-1} - \sum_m^{-1}) \vec{x} + 2 (\sum_m^{-1} M_m - \sum_f^{-1} M_f)^T \vec{x} \quad +$$
$$(M_f^T \textstyle\sum_f^{-1} M_f - M_m^T \sum_m^{-1} M_m + \log(\frac{|\sum_f|}{|\sum_m|})) \qquad = \quad 0.$$

Given a sample point $\vec{x}$ for which the classification is unknown, we evaluate the given formula. If the result is greater than zero, $\vec{x}$ is assigned to the female class. Otherwise, $\vec{x}$ is assigned to the male class.

4.4. Support Vector Machines

Support vector machines can be viewed as an extension of the linear discriminant function. For a detailed description of support vector machines, refer to Cristianini and Shawe-Taylor [8]. The goal is to find a linear discriminant function that separates the data. However, the data is not linearly separable in our case. One way to obtain linearly separable data is to transform the data into a high-dimensional space using a non-linear transformation. For each classification problem that involves two classes of data, a transformation $y : \vec{x} \rightarrow \phi(\vec{x})$ yielding linearly separable classes exists. However, it is difficult to find y. Furthermore, if y was explicitly given, computing the classifier would be computationally inefficient.

To overcome these problems, the classifier is computed without first computing the unknown transformation y. The key property of the function y is that $\langle \phi(\vec{x_i}), \phi(\vec{x_j}) \rangle$ can be rewritten as a kernel function $K(\vec{x_i}, \vec{x_j})$. Using this kernel function, we can compute the separating hyperplane that maximizes the minimum distance
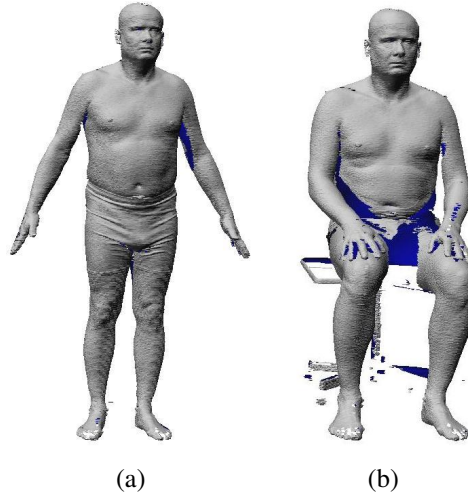

(a)   (b)

Figure 3. *Two postures used for classification.*

between the hyperplane and the closest data point. The solution is the following variation of Equation (7)

$$f(\vec{x}) = \sum_{i \in S} y_i \vec{\alpha}_i K(\vec{x}, \vec{x}_i) + b. \qquad (8)$$

Recall that S is the set of support vector indices and that b is the axis intercept of the hyperplane.

We use the popular radial basis function kernel in our experiments. The best parameter for the kernel is found using grid search. We use the software mySVM [13] in the implementation of this method.

5. Results

This section shows results that demonstrate the accuracy of the proposed classifiers. To improve the classification accuracy, the training data is diagonalized using the learned Gaussian distributions obtained by the maximum likelihood estimate as outlined in Section 4.1. The quality of the classification is measured using the 500 female and 500 male test subjects. As we know the true classes of the 1000 test samples, we can compute the error made by the classifier. The quality of the classification is measured using the 1000 test samples projected into the learned diagonalized space.

First, we discuss the results for training and testing the classifiers using subjects in the same posture as the subject shown in Figure 1. The results of applying the three classifiers reviewed in the previous section are summarized in the first row of Table 1. Support vector machines yield the best result, followed by the linear classifier.

Second, we discuss the results for training the classifiers using subjects in the posture shown in Figure 3(a) and for testing the classifiers using subjects in the posture shown in Figure 3(b).

We test the classifiers using two data sets. First, we test the classifiers using the same subjects that were used for training given in a different posture. Second, we test the classifier using different subjects in a different posture than the ones used for training. It is expected that the first test yields a higher classification accuracy than the second test.

Accuracy	Linear Discriminant	Bayesian Discriminant	Support Vector Machines
Same Posture	94.8	94.7	95.1
Different Posture Same Subjects	94.1	80.8	95.0
Different Posture Different Subjects	92.2	79.7	93.0
Train and Test Different Postures	92.6	92.9	96.1

Table 1. *Classification accuracy of the diagonalized data sets. All the accuracies are given in* %.

We use the same classifiers as in the previous section. The second and third rows of Table 1 give the classification accuracy for the two experiments. As before, support vector machines yield the highest accuracy, followed by the linear classifier. When using different postures, using the same subjects in different postures yields a higher accuracy than using different subjects. This is to be expected as the geodesic distances on one subject are nearly posture-invariant if we neglect topological changes.

Figure 4 shows the test data set using different subjects in different postures than the training data set projected to a plane. Female data projections are shown in red and male data projections are shown in blue. For some of the data points, the body shape is shown. We can see that in this projection, the data is well separated.

Finally, we use 500 subjects in both postures shown in Figure 3 for training and 500 subjects in both postures shown in Figure 3 for testing. We treat two postures of the same subject as independent subjects. That is, when testing the classifier, it is possible that the first posture of a subject is classified correctly while the second posture of the same subject is classified incorrectly.

We first reduce the dimensionality of the data as outlined in Section 3. As a different data set is used for training the classifier, a different set of features are found to be useful for classification. In this experiment, only the four components shown in Figure 5 were found relevant according to Formula 1. Note that those components strongly overlap with the ones found in Section 3. The main difference is that the distances from the chest to the wrists no longer are useful for classification. That is, only distances from the lower back to the face are considered for classification when both postures are used to train the classifier.

The last row of Table 1 gives the classification accuracy. As before, the support vector machine classifier yields the highest accuracy. In this experiment, the Bayesian classifier yields a higher classification accuracy than the linear classifier. The accuracy of all the classifiers is high given that only four features are used for discrimination.

In all of the experiments, the support vector machine classifier yields the best result. All of the results have high classification accuracy. This shows that geodesic distances are suitable to discriminate humans by gender.

6. Conclusion and Future Work

This paper presented a new technique for posture-invariant gender classification of human bodies. Classifiers to distinguish between male and female bodies were learned using a training data
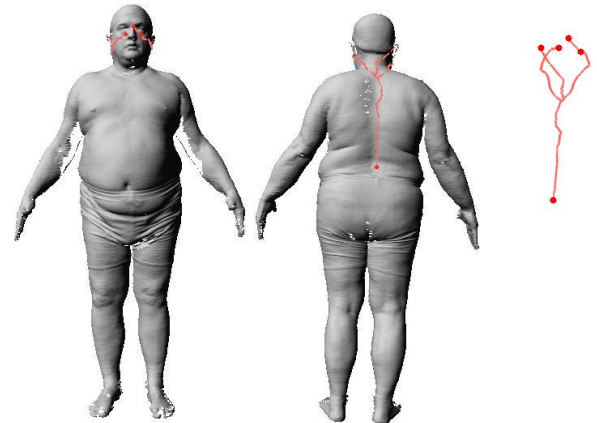

Figure 5. *Subgraph used for classification shown on one of the CAESAR models.*

set and their accuracy was analyzed using a test data set. Experimental results demonstrated the applicability of this method for posture-invariant classification by showing that as few as four features yield a classification accuracy of at least 93%.

The classification is based on the most suitable geodesic distances between discrete landmark positions. We conclude that the most important geodesic distances for gender classification connect the lower back to the face and the chest to the wrists.

An interesting consideration for future work is to use more densely distributed markers as landmarks. To achieve this, markers need to be automatically computed on all of the given subjects.

References

[1] N. R. P. 1024. *Anthropometric Source Book - Volume 1*. NASA, 1978.

[2] N. R. P. 1024. *Anthropometric Source Book - Volume 2*. NASA, 1978.

[3] N. R. P. 1024. *Anthropometric Source Book - Volume 3*. NASA, 1978.

[4] Z. B. Azouz, P. Bose, C. Shu, and S. Wuhrer. Approximations of geodesic distances for incomplete triangular manifolds. In *Proceedings of the 19th Canadian Conference on Computational Geometry*, pages 177–180, 2007.

[5] Z. B. Azouz, C. Shu, and A. Mantel. Automatic locating of anthropometric landmarks on 3d human models. In *International Symposium on 3D Data Processing, Visualization and Transmission*, 2006.
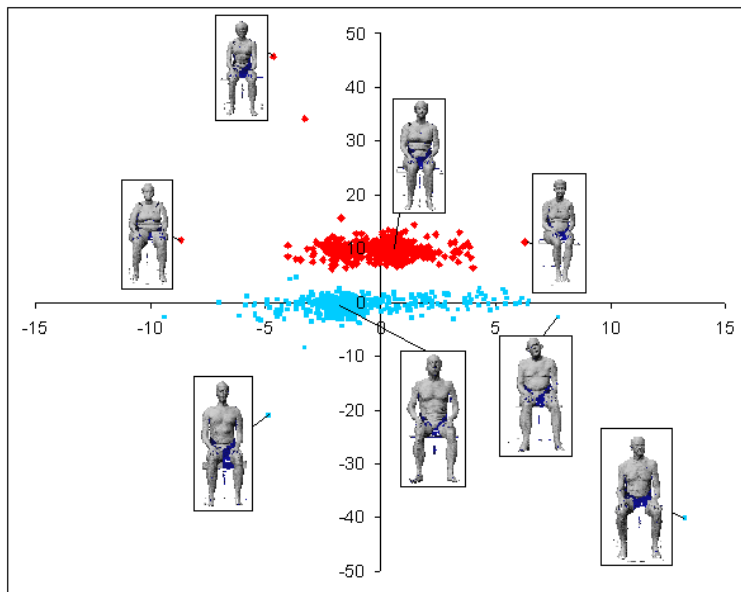
Figure 4. *Projection of the test data sets to a plane.*

[6] D. Burnsides and K. Robinette. 3-d landmark detection and identification in the caesar project. In *International Conference on 3-D Digital Imaging and Modeling*, pages 393–298, 2001.

[7] M. Cerney and D. Adams. Sequestering size: The role of allometry and gender in digital human modeling. *SAE Transactions Journal of Aerospace*, 113:208–214.

[8] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-Related Methods*. Cambridge University Press, 2000.

[9] F. de Helguero. Sui massimi delle curve dimorfiche. *Biometrika*, 3(1):84–98.

[10] R. Duda, P. Hart, and D. Stork. *Pattern Classification, Second Edition*. John Wiley & Sons, Inc., 2001.

[11] A. Godil, P. Grother, and S. Ressler. Human identification from body shape. In *International Conference on 3-D Digital Imaging and Modeling*, pages 386–392, 2003.

[12] A. B. A. Graf and F. A. Wichmann. Gender classification of human faces. In *Proceedings of the Second International Workshop on Biologically Motivated Computer Vision*, pages 491–500, 2002.

[13] T. Joachims. *Making large-Scale SVM Learning Practical*. In: Advances in Kernel Methods - Support Vector Learning, Chapter 11. MIT Press, 1999.

[14] R. Kimmel and J. Sethian. Computing geodesic paths on manifolds. *Proceedings of the National Academy of Sciences*, 95:8431–8435, 1998.

[15] L. Lee and E. Grimson. Gait analysis for recognition and classification. In *International Conference on Automatic Face and Gesture Recognition*, pages 148–155, 2002.

[16] E. Mäkinen and R. Raisamo. An experimental comparison of gender classification methods. *Pattern Recognition Letters*, 29:1544–1556, 2008.

[17] A. O'Toole, K. Deffenbacher, D. Valentin, K. McKee, D. Huff, and H. Abdi. The perception of face gender: the role of stimulus structure in recognition and classification. *Memory & Cognition*, 26(1), 1998.

[18] M. Schilling, A. Watkins, and W. Watkins. Is human height bimodal? *The American Statistician*, 56(3):223–229, 2002.

[19] V. Surazhsky, T. Surazhsky, D. Kirsanov, S. J. Gortler, and H. Hoppe. Fast exact and approximate geodesics on meshes. *ACM Transactions on Graphics*, 24(3):553–560, 2005. Proceedings of SIGGRAPH.