

Gender Recognition from Walking Movements using Adaptive Three-Mode PCA*

James W. Davis Hui Gao
Dept. of Computer Science and Engineering
Ohio State University
Columbus, OH 43210 USA
{jwdavis, gaoh}@cse.ohio-state.edu

Abstract

We present an adaptive three-mode PCA framework for recognizing gender from walking movements. Prototype female and male walkers are initially decomposed into a sub-space of their three-mode components (posture, time, gender). We then assign an importance weight to each motion trajectory in the sub-space and have the model automatically learn the weight values (key features) from labeled training data. We present experiments of recognizing physical (actual) and perceived (from perceptual experiments) gender for 40 walkers. The model demonstrates greater than 90% recognition for both contexts and shows greater flexibility than standard PCA.

1. Introduction

Human movement patterns contain subtle, yet informative, stylistic variations. Even from point-light displays, people can recognize the gender of a walker [16, 1, 8, 20], identify individuals from their gait [7, 2], and estimate dynamics from movement (e.g., estimating the weight of an object being lifted) [21, 3]. Our overall goal is to develop a computational framework to quantitatively analyze, model, and recognize human action styles from the underlying movement patterns. In this paper we focus on the task of recognizing gender from walking sequences. Our approach factorizes the motion trajectories of gender-prototype walkers into three-mode PCA components (posture, time, and gender), which enables us to employ an importance weight on each trajectory to bias the recognition process using the most informative (key) features.

The use of trainable feature weights enables gender recognition under different criteria. Do we wish to model/recognize the *real* (actual) gender or the *perceived* maleness or femaleness of the walker? Although men and women tend to have different gait patterns, there are some

individuals whose walking styles visually coincide with the opposite gender [16]. Hence, the physical and perceptual gender labels may be in conflict. For automatic visual surveillance, recognizing the *physical* gender is of most concern, whereas a model of the *perceptual* gender appearance is most important to computer animation or for search-by-example in motion libraries. Understanding the perceived gender is also important for studying which key features humans consistently use to discriminate gender. We address both contexts with a single trainable framework that learns the set of key features for each context.

The proposed three-mode PCA framework is applicable to other style variations besides gender. In previous work, we examined the stylistic movement changes due to carrying load and walking pace [12, 10, 11]. A brief result on physical gender recognition is also presented in [10]. In this paper, we describe a variant of the three-mode approach for two styles (gender) and concentrate fully on its application to the gender recognition task. We also present perceptual gender recognition results (using perceptual experiments on human subjects and our computational model) as well as physical gender results.

In the remainder of the paper, we first discuss related work in Sect. 2. We then present our expressive three-mode framework in Sect. 3, including an outline of the general three-mode factorization technique. In Sect. 4, we examine our model for both physical and perceptual gender recognition contexts with a large set of 40 walkers. Lastly, we conclude with a summary of the research in Sect. 5.

2. Related Work

In computer vision, action style approaches include modeling spatial pointing gestures by a Parameterized-HMM [27], discriminating children from adults by relative stride information [9], classifying typical from atypical gaits by considering motion regularities [14], recognizing atomic activities by PCA and analytical global transformations [28], classifying human ID and gender from gait using

*Appears in *IEEE Workshop on Articulated and Nonrigid Motion*, Washington DC, June 27, 2004.

SVMs [19], and representing complex motion patterns by linear combinations of prototype sequences [15].

In computer animation, a Fourier-based approach was used in [24] to generate human motion with emotional properties. An HMM with entropy minimization was used by [4] to generate different state-based animation styles. A factorization of motion-capture data for extracting person-specific motion signatures was described in [25]. A movement exaggeration model using measurements of the observability and predictability of joint angle trajectories was presented in [13]. In [6], the EMOTE character animation system using LMA effort and shape components was employed to generate natural synthetic gestures.

Regarding gender recognition, past perceptual experiments with human observers have shown average gender recognition rates for adult walkers within 63-76% [16, 1, 22]. Much of that work has focussed on determining the key perceptual features. Static features (e.g., shoulder-hip ratio or center-of-moment [1, 8]) as well as dynamic features (e.g., lateral body sway [20]) have been proposed. In [16] it was reported that one female walker was identified as male in most of the trials, illustrating that physical and perceptual gender do not necessarily coincide.

A two-stage PCA framework for recognizing gender was recently proposed in [22]. The first PCA decomposed each walker’s data into its Eigenspace, and a second PCA was applied to all walker Eigenspaces followed by a linear gender classifier. In comparison to our method, we use a single three-mode PCA factorization to separate the movements of prototype female and male walkers into posture, time, and gender basis sets. This representation enables us to embed adaptable weights on each motion trajectory in the sub-space to denote the importance of each trajectory when estimating the gender. Our approach can be easily adapted to training data with different gender recognition contexts (actual vs. perceived), and hence is more flexible than standard/fixed PCA recognition. This issue of gender contexts has not been explored in past approaches.

3. Expressive Three-Mode PCA Model

Human movements in the most basic sense can be described as a body posture changing over time (two modes: posture, time). When considering stylistic movement variations, due to gender, we have a total of three modes to describe the action: posture, time, and *gender*. We exploit the tri-modal nature of stylistic motion trajectories with an efficient three-mode PCA factorization.

3.1. Three-Mode PCA

As an extension of the traditional two-mode PCA, three-mode factorization [23] produces three orthonormal basis

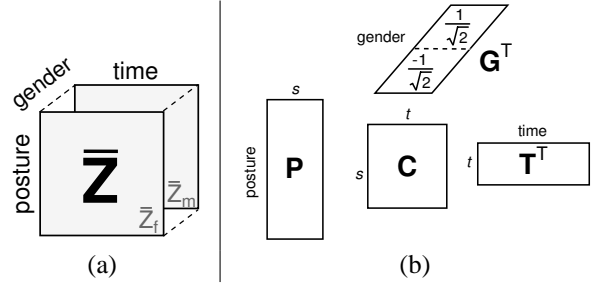


Figure 1: (a) Three-mode arrangement of two gender prototypes. (b) Three-mode factorization of gender prototypes.

sets for motion trajectories arranged in a 3-D cube. For gender recognition, the style dimension corresponds to a binary gender assignment of FEMALE or MALE. We form a cube \bar{Z} using two prototype walkers constructed by averaging the walking examples in each gender class. The walking data are first pre-aligned (see Sect. 4) using cycle extraction (at the same phase), height normalization, and time normalization to N frames.

For M trajectories of each walker, each gender prototype is represented as a matrix of size $M \times N$. We place the two prototypes into the first and second (last) frontal plane of the cube, and center the cube (mean-subtract) along the gender mode. The result is a gender cube \bar{Z} of size $M \times N \times 2$ for the mean-subtracted female and male prototypes \bar{Z}_f, \bar{Z}_m (see Fig. 1.a).

A three-mode factorization of \bar{Z} produces three orthonormal basis sets P, T , and G that span column (posture), row (time), and slice (gender) dimensions (see Fig. 1.b). The reconstruction of \bar{Z} (flattened) can be written as

$$[\bar{Z}_f | \bar{Z}_m] = PC(G^T \otimes T^T) \quad (1)$$

where \otimes is the Kronecker product and C is the core matrix [17]. The three basis sets can be solved from three different flattening arrangements of \bar{Z}

$$\text{Posture: } P = \text{colSpace}([\bar{Z}_f | \bar{Z}_m]) \quad (2)$$

$$\text{Time: } T = \text{colSpace}([\bar{Z}_f^T | \bar{Z}_m^T]) \quad (3)$$

$$\text{Gender: } G = \text{rowSpace}([\bar{Z}_f | \bar{Z}_m]) \quad (4)$$

where $\bar{Z}_{\{f,m\}}^T$ is the transpose of $\bar{Z}_{\{f,m\}}$, and $\bar{Z}_{\{f,m\}}$ is the rasterized column vector of matrix $\bar{Z}_{\{f,m\}}$ (concatenation of trajectories into a single column vector). The desired column and row spaces in Eqns. 2-4 can be found using Singular Value Decomposition (SVD). Note that only a few components in P and T are typically needed to capture most of the variance in the data, and that no two of the three basis sets can be produced within a single two-mode (matrix) SVD factorization of a flattened \bar{Z} .

Using the two gender prototypes, \bar{Z}_f and \bar{Z}_m , the normalized gender basis is constrained to be $G = [-1, 1]^\top / \sqrt{2}$, signifying the female ($\frac{-1}{\sqrt{2}}$) and male ($\frac{1}{\sqrt{2}}$) genders. The core matrix C can be solved by re-arranging Eqn. 1 as

$$C = P^\top [\bar{Z}_f | \bar{Z}_m] (G^\top \otimes T^\top)^\top \quad (5)$$

Related methods for solving the three-mode factorization can be found in [18, 26].

3.1.1. Prototype Reconstruction

The three-mode formulation in Eqn. 1 can also be used to individually reconstruct the gender prototypes \bar{Z}_f and \bar{Z}_m with

$$\bar{Z}_{\{f,m\}} = PCg_{\{f,m\}}T^\top \quad (6)$$

where the gender parameter $g_{\{f,m\}} = \frac{\pm 1}{\sqrt{2}}$ signifies the gender ($\frac{-1}{\sqrt{2}}$ for female, $\frac{1}{\sqrt{2}}$ for male). Rewriting Eqn. 6 as a summation of three-mode basis elements, we can isolate the gender parameter from the remaining factored terms

$$\bar{Z}_{ij\{f,m\}} = \sum_{p=1}^s \sum_{q=1}^t P_{ip} C_{pq} g_{\{f,m\}} T_{jq} \quad (7)$$

$$= g_{\{f,m\}} \left(\sum_{p=1}^s \sum_{q=1}^t P_{ip} C_{pq} T_{jq} \right) \quad (8)$$

$$= g_{\{f,m\}} \cdot \alpha_{ij} \quad (9)$$

where the indices i, j correspond to elements in the respective posture and time dimensions ($1 \leq i \leq M$, and $1 \leq j \leq N$).

3.1.2. Gender Parameter Estimation

Following Eqn. 9, the unknown gender parameter \hat{g} for a new walker \hat{z} (after subtracting the model mean) can be estimated by finding the value of \hat{g} that minimizes the sum-of-squared-error (SSE) reconstruction error

$$\mathcal{F} = \sum_{i=1}^M \sum_{j=1}^N (\hat{z}_{ij} - \hat{g} \cdot \alpha_{ij})^2 \quad (10)$$

Setting the partial derivatives of \mathcal{F} to zero and re-arranging the equation, the resulting gender parameter \hat{g} is given by

$$\hat{g} = \frac{\sum_i \sum_j \hat{z}_{ij} \cdot \alpha_{ij}}{\sum_i \sum_j \alpha_{ij}^2} \quad (11)$$

where essentially the gender parameter is computed by projecting \hat{z} onto the normalized α_{ij} basis elements. The final gender label can be assigned by examining the sign of

\hat{g} , choosing FEMALE if it is negative and MALE if positive (i.e., selecting the nearest centered gender prototype).

The above gender estimation in Eqn. 11 could have equivalently been achieved by rasterizing the gender prototype data into a $MN \times 2$ matrix and performing a standard two-mode PCA to estimate the gender parameter. We will show that the prototype three-mode formulation (Eqn. 11) will enable us to easily embed adaptable weights *within* the PCA representation to influence the direct computation of the gender parameter (the projection coefficient). Alternatively, all walking examples could be employed in a rasterized two-mode PCA and have weights applied to the basis vectors, but these weights would generally not have any correspondence back to the original (and distinct) motion trajectories in which certain key features may reside.

3.2. Extended Three-Mode PCA

The SSE minimization in Eqn.10 gives significant influence to trajectories having large magnitude differences from the model-mean. But perhaps only certain trajectories carry the true expressive information regarding the gender differences. Furthermore these trajectories could have smaller magnitude differences, which unfortunately would be attenuated in a standard SSE estimation.

To handle these problems, we introduce a weighting factor on each trajectory to denote its impact for estimating the context-based gender assignments. We update Eqn. 10 as

$$\mathcal{F} = \sum_{i=1}^M \mathcal{E}_i \sum_{j=1}^N (\hat{z}_{ij} - \hat{g} \cdot \alpha_{ij})^2 \quad (12)$$

with a positive weight \mathcal{E}_i on each of the M trajectories. The gender parameter can then be estimated using

$$\hat{g} = \frac{\sum_i \mathcal{E}_i \sum_j \hat{z}_{ij} \cdot \alpha_{ij}}{\sum_i \mathcal{E}_i \sum_j \alpha_{ij}^2} \quad (13)$$

$$= \frac{\sum_i \mathcal{E}_i \Delta_i}{\sum_i \mathcal{E}_i \sum_j \alpha_{ij}^2} \quad (14)$$

$$= \sum_{i=1}^M \tilde{\mathcal{E}}_i \Delta_i \quad (15)$$

where $\Delta_i = \sum_{j=1}^N \hat{z}_{ij} \cdot \alpha_{ij}$. As the denominator $\sum_i \mathcal{E}_i \sum_j \alpha_{ij}^2$ in Eqn. 14 is a constant for a given set of \mathcal{E}_i , we fold this term into the ‘‘expressive weights’’ $\tilde{\mathcal{E}}_i$ in Eqn. 15. Note that if we set each expressive weight to $(\sum_i \sum_j \alpha_{ij}^2)^{-1}$ in Eqn. 15, the resulting gender parameter estimation reverts to the previous SSE method (Eqn. 11) which is equivalent to standard PCA. We next need to learn the appropriate values for the expressive weights.

3.2.1. Learning Expressive Weights

Our approach to learning the expressive weights $\tilde{\mathcal{E}}_i$ is to minimize a second error function that compares the computed gender parameters \hat{g} (using Eqn. 15) with pre-assigned gender parameter values ($\bar{g} = \frac{\pm 1}{\sqrt{2}}$) for K different training examples

$$J = \sum_{k=1}^K (\bar{g}_k - \hat{g}_k)^2 \quad (16)$$

$$= \sum_{k=1}^K (\bar{g}_k - \sum_{i=1}^M \tilde{\mathcal{E}}_i \Delta_{ik})^2 \quad (17)$$

The expressive weights in Eqn. 17 can be solved linearly, but this requires as many training examples as the number of expressive weights ($K \geq M$). It would also be difficult to constrain the weights to be positive. Instead, we employ a fast iterative gradient descent algorithm [5] of the form

$$\tilde{\mathcal{E}}_i(n+1) = \tilde{\mathcal{E}}_i(n) - \eta(n) \cdot \frac{\partial J}{\partial \tilde{\mathcal{E}}_i} \quad (18)$$

with the gradients $\frac{\partial J}{\partial \tilde{\mathcal{E}}_i}$ computed as

$$\frac{\partial J}{\partial \tilde{\mathcal{E}}_i} = -2 \sum_{k=1}^K \Delta_{ik} (\bar{g}_k - \sum_{z=1}^M \tilde{\mathcal{E}}_z \Delta_{zk}) \quad (19)$$

The learning rate η is re-computed at each iteration (via interpolation of the error function) to yield the best incremental update. The expressive weights are initialized to the default SSE formulation, $\tilde{\mathcal{E}}_i(0) = (\sum_i \sum_j \alpha_{ij}^2)^{-1}$, and are confined to be positive in each iteration.

Following the termination of Eqn. 18, the gender parameter for a walker can be estimated with Eqn. 15 using the newly learned expressive weights. As before, the sign function can be used to threshold the estimated gender parameter \hat{g} to compute a final gender assignment (-1 = FEMALE, +1 = MALE). With non-uniform values for the expressive weights $\tilde{\mathcal{E}}_i$, the approach is capable of emphasizing select trajectories and producing non-SSE gender parameter estimations according to a specific recognition context.

4. Experiments

For our experiments, we first examined the gender recognition capability of human observers with our data set. Then we tested our adaptable PCA model for recognizing gender in both physical and perceptual contexts.

To focus on the general mechanism for gender *recognition* (rather than extracting trajectories directly from video), our data set was collected using motion-capture. Gender recognition has widely been, and continues to be, studied

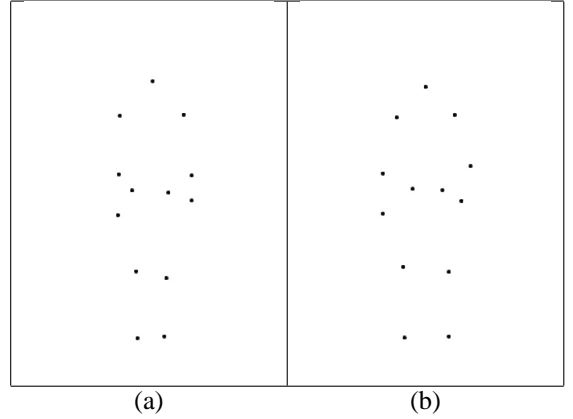


Figure 2: Point-light images of a (a) female walker and (b) male walker.

using point-light trajectories. Though we employ motion-capture trajectories in our experiments, our general learning and recognition approaches are an important aspect of computer vision for motion understanding. In future work, we plan to incorporate a video-based human body tracker as input.

Our data set includes 40 adult walkers (20 female and 20 male) collected by N. Troje at the BioMotionLab of the Ruhr-University in Bochum, Germany, where ten gait cycles of each walker were recorded with a Vicon motion-capture system. We used 13 point-lights for each walker, where 10 were markers located closely to the major limb joints and 3 were virtual points created by averaging four head markers (into 1 head point) and two hip markers on the left and right side (into 2 hip markers).

Each walking sequence was low-pass filtered (cutoff 6 Hz) and rotated to face the forward direction. For each point-light walker, we normalized the height using its average tibia length. We then extracted one walk cycle from each sequence (at the same phase) by detecting cyclic curvature peaks in the left-knee trajectory, and time-normalized each cycle to $N = 50$ frames (using spline interpolation) to remove any influence of walking speed. Any discontinuity between the last and first frame for each trajectory $x(t)$ was removed by distributing the error $\delta = x(1) - x(N)$ throughout the trajectory as $\tilde{x}(t) = x(t) + (t-1) \cdot \delta/N$, $1 \leq t \leq N$. Lastly, the 3-D trajectories were orthographically projected into 2-D at the frontal view (the best gender discrimination view [22]). Example point-light images are shown in Fig. 2.

4.1. Perceptual Recognition

A computer program was implemented to collect human judgments of the gender for the 40 walkers. Each point-light walker (in random order) was presented on a 21 in.

CRT monitor (1280 × 1024) and looped until the observer selected a gender label using the keyboard (‘F’ for FEMALE or ‘M’ for MALE). The participant was seated at a typical working distance, and no time restriction was enforced.

Each 50-frame walking cycle was rendered as black dots against an off-white background at 36 FPS, corresponding to a cycle time of 1.4s (maximum cycle time for all 40 walkers). The height of each walker was scaled to 70% of the screen resolution height. Additionally, the root of each walker (center of two hip markers) was randomly positioned within a small circle at the center of screen (radius of 10% the screen resolution height) to prevent any explicit spatial comparison.

Fifteen university students (5 female, 10 male) were recruited as participants for the experiment. The average gender recognition rate was 69%, which is within the range of previously reported results [16, 1, 22].

We additionally computed a *gender consistency* value for each walker. The gender selection for each walker (from each participant) was assigned a value of -1 for FEMALE and $+1$ for MALE. We computed the gender consistencies by averaging the ± 1 values assigned to each walker by the 15 participants. A consistency value of -1 corresponds to total agreement of the walker as FEMALE, a value of $+1$ corresponds to total agreement as MALE, and a value near zero corresponds to AMBIGUOUS. We present the resulting gender consistencies for the 40 walkers in Fig. 3.

Three walkers (#1, #4, #34) were unanimously labeled by all 15 observers. Interestingly, one of them (#4) was a female that was perceptually labeled as male by all participants. This clearly demonstrates the potential differences between the perceived and actual gender. Several other walkers were difficult to label (consistency values near 0).

4.2. Three-Mode PCA: Physical Gender

In this experiment, we examined our expressive three-mode approach for recognizing the physical gender of the walkers. For training, we assigned a gender parameter value of $\frac{-1}{\sqrt{2}}$ to the *true* females and $\frac{1}{\sqrt{2}}$ to the *true* males.

To avoid overfitting the model, we employed a leave-one-out cross-validation technique on the walkers. We varied the percent modal fit for the basis sets P and T (using the least number of components in each basis to capture the desired percent of variance) of the prototypes (computed from 39 training examples). The optimal expressive weights (see Fig. 4) were computed by averaging the 40 sets of weights for the modal fit at 75%, which had the lowest cross-validation testing error of 25%. The ordering of the 26 weights correspond to {HEAD:(1-2), RIGHT-ARM:(3-8), LEFT-ARM:(9-14), RIGHT/LEFT-HIP:(15-18), RIGHT-LEG:(19-22), and LEFT-LEG:(23-26)}. Some weights were zero, signifying that they were not key features.

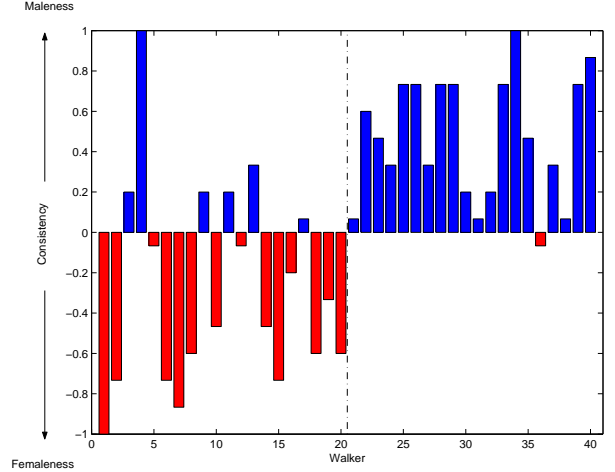


Figure 3: Gender consistency values for the 40 walking sequences (females: #1-20, males: #21-40)

Next we computed two prototypes from all 40 walkers and constructed a single three-mode factorization at the selected modal fit. The resulting three-mode basis captured 98% of the variance in the gender prototypes. The basis sets P and T were of dimension 26×3 and 50×3 , respectively (the core C was therefore of size 3×3).

To evaluate this model with the average expressive weights, we first computed the gender parameters for all 40 walkers using Eqn. 15, and compared them to the $\frac{\pm 1}{\sqrt{2}}$ physical gender values assigned to the training data (see Fig. 5.a). Thresholding the gender parameter values using the sign function produced a 92.5% classification rate.

For comparison, we employed the same cross-validation technique on a three-mode model without any expressive weights (i.e., the default SSE estimation). The best modal fit occurred at a 50% fit with a testing error of 30%. This resulted in a much noisier gender parameter estimation for the overall model of 40 walkers (see Fig. 5.b) and yielded a much lower 70% classification rate. The gender estimations with our expressive model appear much closer to the desired gender values (average difference = .31) than the alternative SSE estimations (average difference = 1.07).

4.3. Three-Mode PCA: Perceptual Gender

An advantage to our framework is that the model can easily adapt to differently-labeled training data (different contexts). To demonstrate this capability, we trained our model to produce gender estimations similar to the classification results produced by the human observers.

We first thresholded the gender consistency values from the perceptual experiment (Sect. 4.1) to label the walkers with their perceived gender. For each walker in the training

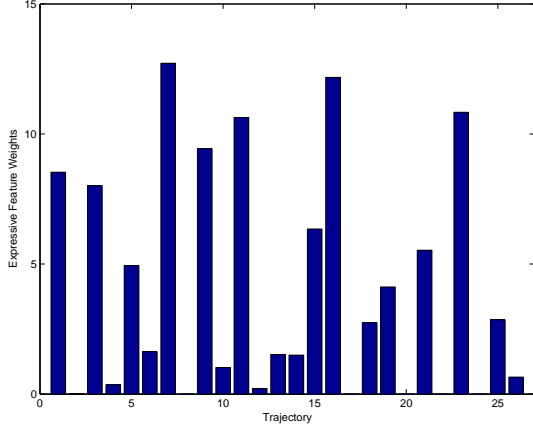
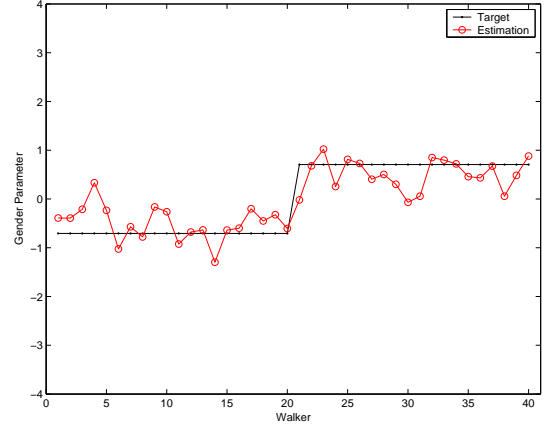


Figure 4: Average expressive weights for physical gender.

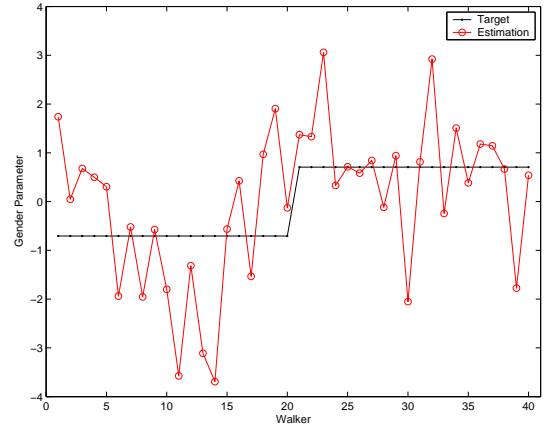
set, we assigned a perceptual gender label of $\frac{-1}{\sqrt{2}}$ (FEMALE) if it had a negative gender consistency or $\frac{1}{\sqrt{2}}$ (MALE) otherwise. The perceptual training set resulted in 15 “perceived females” (14 true females, 1 true male) and 25 “perceived males” (19 true males, 6 true females).

We used the same cross-validation technique outlined in Sect. 4.2 to construct the optimal expressive PCA model. The two prototypes were constructed using the perceived-female and perceived-male examples (not the true gender). The expressive weights (see Fig. 6) were generated by averaging the 40 sets of expressive weights computed at the cross-validation modal fit of 80%, which had the smallest testing error of 27.5%. The three-mode basis (using all examples for the prototypes) captured 98% of the variance in the two gender prototypes. The basis sets P and T were of dimension 26×4 and 50×4 , respectively (the core C was of size 4×4).

The gender parameter results for the expressive model are shown with the target values in Fig. 7.a. Thresholding the gender parameter values using the sign function produced a 90% classification rate. For comparison, we again computed the optimal three-mode model without any weights (SSE estimation) for the perceptual data using the cross-validation technique. The best modal fit occurred at 50% with a testing error of 42.5%. The results showed noisier gender estimations (see Fig. 7.b). The thresholded gender parameters produced a lower 77.5% classification rate. As in the previous physical gender context, our expressive model produced gender parameter values much closer to the desired perceptual values (average difference = .33) than did the alternative SSE-based approach (average difference = .68).



(a)



(b)

Figure 5: Physical gender estimation. (a) Expressive three-mode. (b) Non-expressive SSE three-mode. Walkers 1-20:female, 21-40: male.

4.3.1. Consistency Weighting

To account for the gender ambiguity that occurs for some walkers (having gender consistency values near zero), we can attenuate the influence of those walkers during the learning phase for the expressive weights to give more emphasis to the remaining walkers with high consistency magnitudes. Given the set of K training walkers and their assigned perceptual genders \bar{g}_k , we slightly alter the previous matching error function (Eqn. 17) by using their consistency magnitudes ω_k to bias the minimization procedure to those examples having more reliable gender assignments across the observers

$$J_p = \sum_{k=1}^K \omega_k \cdot \left(\bar{g}_k - \sum_{i=1}^M \tilde{\mathcal{E}}_i \Delta_{ik} \right)^2 \quad (20)$$

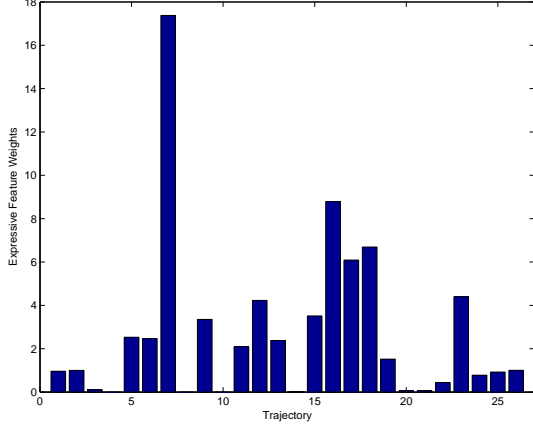


Figure 6: Average expressive weights for perceptual gender.

The corresponding perceptual gradient is then

$$\frac{\partial J_p}{\partial \tilde{\mathcal{E}}_i} = -2 \sum_{k=1}^K \omega_k \Delta_{ik} (\bar{g}_k - \sum_{z=1}^M \tilde{\mathcal{E}}_z \Delta_{zk}) \quad (21)$$

This new gradient is used as before in the gradient descent procedure (Eqn. 18) to determine the appropriate expressive weights for the perceptually-labeled walkers.

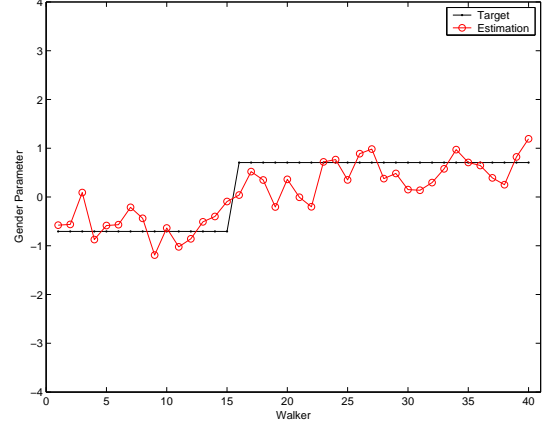
With this new approach, the perceptually ambiguous walkers will be mostly disregarded when learning the expressive weights. Therefore, an error for a highly ambiguous walker should not be as equally counted as the other consistently-labeled walkers. We correspondingly modify the standard recognition error rate to now be weighted by the consistency magnitudes

$$Error = \frac{\sum_k \omega_k [Sgn(\bar{g}_k) \neq Sgn(\hat{g}_k)]}{\sum_k \omega_k} \quad (22)$$

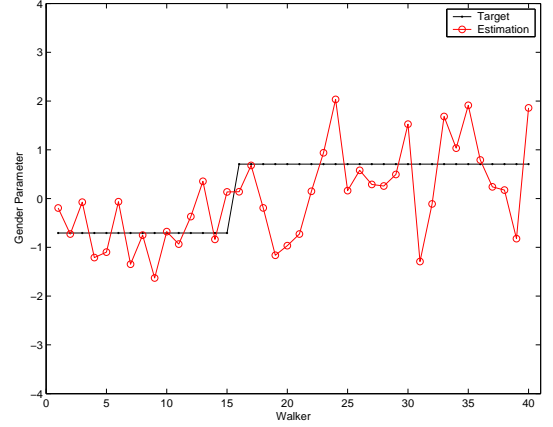
Using the new perceptual gradient function and weighted error calculation in the cross-validation technique, the single optimal expressive model was computed and tested on the walking data. The resulting expressive model produced the weighted classification rate of 95.5% for all 40 walkers (the corresponding optimal SSE model produced a weighted classification rate of 88%).

5. Summary and Conclusion

We presented a computational approach for recognizing the gender of walking people using an efficient and adaptable three-mode PCA framework. The approach initially factorizes prototype female and male walkers into their three-mode basis sets representing the body posture, temporal trajectories, and gender changes. This multi-modal decomposition of the data is suitable for incorporating expressive



(a)



(b)

Figure 7: Perceptual gender estimation. (a) Expressive three-mode. (b) Non-expressive SSE three-mode. Walkers #1-15: perceived-female, #16-40: perceived-male.

weights on motion trajectories to bias the model estimation of gender. The method embeds an expressive weight on each trajectory in the sub-space and automatically learns the necessary weight values from gender-labeled training data. One advantage of this approach is that it can adapt to different recognition contexts (physical and perceptual gender) for the same underlying data. This concept has not been addressed in the gender recognition literature.

Our expressive three-mode PCA model produced a recognition rate of 92.5% for 40 walkers (20 female, 20 male) labeled with their physical (true) gender. A gender classification experiment using 15 participants resulted in a 69% perceptual classification rate (within the range of previously reported results). Trained using the dominant perceptual label for each walker, the expressive PCA model produced a 90% recognition rate. An additional model trained with consistency-weighted (perceptual confidence)

examples produced a higher 95.5% weighted classification rate. The method demonstrated that our approach can successfully and automatically adapt to different gender contexts using expressive weights, and that it outperforms a standard three-mode (or rasterized two-mode) PCA model with no expressive weights.

Gender recognition has been an active research domain for several years. We proposed a new computational approach that could potentially provide insight into the features most useful for classification (perceptual and computational). In future work we plan to explore this possibility further by examining different feature representations (other than motion trajectories) for multi-phase (non-periodic) activities. We also intend to incorporate a video-based human body tracker and apply the three-mode PCA framework to other recognition (video surveillance) and synthesis (computer animation) tasks.

Acknowledgments

This research was supported in part by the National Science Foundation under grant No. 0236653 and OBR Hayes Doctoral Incentive Fund Grant Program Fellowship. We additionally thank N. Troje at the BioMotionLab of the Ruhr-University in Bochum, Germany for supplying the motion-capture walking data.

References

- [1] C. Barclay, J. Cutting, and L. Kozlowski. Temporal and spatial factors in gait perception that influence gender recognition. *Perception & Psychophysics*, 23(2):145–152, 1978.
- [2] T. Beardsworth and T. Buckner. The ability to recognize oneself from a video recording of one’s movements without seeing one’s body. *Bull. Psych. Soc.*, 18:19–22, 1981.
- [3] G. Bingham. Kinematic form and scaling: Further investigations on the visual perception of lifted weight. *J. of Exp. Psych.*, 13(2):155–177, 1987.
- [4] M. Brand and A. Hertzmann. Style machines. In *Proc. SIGGRAPH*, pages 183–192. ACM, July 2000.
- [5] R. Burden and J. Faires. *Numerical Analysis*. PWS, Boston, 1993.
- [6] D. Chi, M. Costa, L. Zhao, and N. Badler. The EMOTE model for effort and shape. In *Proc. SIGGRAPH*, pages 173–182. ACM, 2000.
- [7] J. Cutting and L. Kozlowski. Recognizing friends by their walk: Gait perception without familiarity cues. *Bull. Psych. Soc.*, 9:353–356, 1977.
- [8] J. Cutting, D. Proffitt, and L. Kozlowski. A biomechanical invariant for gait perception. *J. of Exp. Psych.*, 4(3):357–372, 1978.
- [9] J. Davis. Visual categorization of children and adult walking styles. In *Proc. Int. Conf. Audio- and Video-based Biometric Person Authentication*, pages 295–300, 2001.
- [10] J. Davis and H. Gao. An expressive three-mode principal components model of human action style. *Image and Vision Computing*, 21(11), 2003.
- [11] J. Davis and H. Gao. Recognizing human action efforts: an adaptive three-mode PCA framework. In *Proc. Int. Conf. Comp. Vis.*, pages 1463–1469, 2003.
- [12] J. Davis, H. Gao, and V. Kannappan. A three-mode expressive feature model of action effort. In *Wkshp. on Motion and Video Computing*, pages 139–144. IEEE, 2002.
- [13] J. Davis and V. Kannappan. Expressive features for movement exaggeration. In *SIGGRAPH Conference Abstracts and Applications*, page 182. ACM, 2002.
- [14] J. Davis and S. Taylor. Analysis and recognition of walking movements. In *Proc. Int. Conf. Pat. Rec.*, pages 315–318, 2002.
- [15] M. Giese and T. Poggio. Morphable models for the analysis and synthesis of complex motion patterns. *Int. J. of Comp. Vis.*, 38(1):59–73, 2000.
- [16] L. Kozlowski and J. Cutting. Recognizing the sex of a walker from dynamic point-light display. *Perception & Psychophysics*, 21(6):575–580, 1977.
- [17] P. Kroonenberg. *Three-Mode Principal Component Analysis Theory and Applications*. DSWO Press, Leiden, 1983.
- [18] P. Kroonenberg and J. Leeuw. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 45(1):69–97, 1980.
- [19] L. Lee and W.E.L. Grimson. Gait analysis for recognition and classification. In *Proc. Int. Conf. on Auto. Face and Gesture Recognition*, pages 155–162. IEEE, 2002.
- [20] G. Mather and L. Murdoch. Gender discrimination in biological motion displays based on dynamic cues. *Proc. R. Soc. Lond. B*, 258:273–279, 1994.
- [21] S. Runeson and G. Frykholm. Visual perception of lifted weight. *J. of Exp. Psych.*, 7(4):733–740, 1981.
- [22] N. Troje. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *J. of Vision*, 2(5):371–387, 2002.
- [23] L. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.
- [24] M. Unuma, K. Anjyo, and R. Takeuchi. Fourier principles for emotion-based human figure animation. In *Proc. SIGGRAPH*, pages 91–96. ACM, 1995.
- [25] M. Vasilescu. Human motion signatures for character animation. In *SIGGRAPH Conference Abstracts and Applications*, page 200. ACM, 2001.
- [26] M. Vasilescu and D. Terzopoulos. Multilinear analysis of image ensembles: TensorFaces. In *Proc. European Conf. Comp. Vis.*, pages 447–460, 2002.

- [27] A. Wilson and A. Bobick. Parametric Hidden Markov Models for gesture recognition. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 21(9):884–900, 1999.
- [28] Y. Yacoob and M. Black. Parameterized modeling and recognition of activities. *Computer Vision and Image Understanding*, 73(2):232–247, 1999.