An expressive three-mode principal components model for gender recognition

James W. Davis

Department of Computer and Information Science, Center for Cognitive Science, Ohio State University, Columbus, OH, USA

Hui Gao

Department of Computer and Information Science, Ohio State University, Columbus, OH, USA

We present a three-mode expressive-feature model for recognizing gender (female, male) from point-light displays of walking people. Prototype female and male walkers are initially decomposed into a subspace of their three-mode components (posture, time, and gender). We then apply a weight factor to each point-light trajectory in the basis representation to enable adaptive, context-based gender estimations. The weight values are automatically learned from labeled training data. We present experiments using physical (actual) and perceived (from perceptual experiments) gender labels to train and test the system. Results with 40 walkers demonstrate greater than 90% recognition for both physically and perceptually-labeled training examples. The approach has a greater flexibility over standard squared-error gender estimation to successfully adapt to different matching contexts.

Keywords: gender classification, recognition, three-mode principal components, style analysis, biological motion

Introduction

Popularized by Johansson in the early 1970s (Johansson, 1973), point-light displays provide an ideal means to study the contribution of motion to the perception of biological (and mechanical) movements. In classic experiments examining human movements, the visual details of the human body were hidden except for several bright point-lights located at the major limb joints (shown against a dark background). Although sometimes difficult to interpret from single or multiple static images (Johansson, 1973; Kozlowski & Cutting, 1977), when viewed in a sequence, the moving point-lights convey a vivid and compelling percept of human movement. Observers can further extract many subtle yet informative style properties of the movement. People can recognize gender from gait and facial/head motion (Barclay, Cutting, & Kozlowski, 1978; Cutting, Proffitt, & Kozlowski, 1978; Hill & Johnston, 2001; Kozlowski & Cutting, 1977; Mather & Murdoch, 1994; Troje, 2002), identify individuals from gait patterns and arm movements (Beardsworth & Buckner, 1981; Cutting & Kozlowski, 1977; Hill & Pollick, 2000), infer emotions from dance and arm movements (Brownlow, Dixon, Egbert, & Radcliffe, 1997; Dittrich, Troscianko, Lea, & Morgan, 1996; Pollick, Paterson, Bruderlin, & Stanford, 2001; Walk & Homan, 1984), and estimate dynamics such as the weight of a lifted object (Bingham, 1987; Bingham, 1993; Runeson & Frykholm, 1981; Runeson & Frykholm, 1983).

Regarding gender recognition from point-lights, both human perception and computational algorithms for discriminating female from male walking styles have been explored in the past several years. Results have shown gender recognition performance by humans in the range of 46–86% for different actions, ages, and views (Barclay et al., 1978; Hirashima, 1999; Kozlowski & Cutting, 1977; Montepare & Zebrowitz-McArthur, 1988; Runeson & Frykholm, 1983; Troje, 2002). A recent pattern recognition framework (Troje, 2002) reported a higher 92.5% recognition rate on adult walkers. Much of the effort on gender recognition has focused on the manual identification of key features that enable the perceptual classification between female and male walking styles. Factors related to speed, arm swing, shoulder-hip lengths, inversion, and body sway have been examined. However, to date there is no conclusive evidence as to which features actually drive the discrimination process. No single feature is likely to be sufficient, but rather it seems multiple combined features are involved.

In this work, we evaluate our new expressive three-mode pattern recognition approach (Davis, Gao, & Kannappan, 2002; Davis & Gao, 2003a; Davis & Gao, 2003b) for recognizing the gender of adult point-light walkers. The term "expressive" refers to the framework's ability to weight the representational units differently to best express (capture) the intrinsic style/gender variations (instead of manually assigning the key features). The framework automatically learns the key features for a given representation (e.g., point-light trajectories) by tuning a weight value for each representational unit (e.g., each trajectory) to bias the numerical estimation of class labels to match the target training labels (e.g., FEMALE = –1, MALE = +1).

doi:10.1167/4.5.2 Received September 9, 2003; published May 6, 2004 ISSN 1534-7362

The approach first constructs an efficient principal components analysis (PCA) representation of point-light trajectories for a prototype (average) female and male walker. A large set of gender-labeled point-light walkers are then used to automatically learn which point-light trajectories (and in what combination) in the prototype PCA representation best express the gender of the walkers. The non-expressive trajectories are removed from consideration, and the remaining expressive trajectories are weighted in the PCA representation to bias the gender estimation method to produce the desired gender labels. The approach combines the benefits of PCA and machine learning into a single, robust gender classifier.

The generality of the approach is found in the ability to train the system to recognize different gender contexts of the walkers. The typical gender context is to recognize the true physical gender of the walkers (*is* female/male). An alternate context for the system could be to recognize the "perceived" gender of the walker (*appears* female/male). For example, a female walker could be consistently perceived by several observers to have a male-like gait pattern (appearance vs. truth). Both of these contexts may have particular applied relevance. For automatic visual surveillance, recognizing the physical gender is of most concern, whereas a model of the perceived gender is most important for computer animation tools to give the best appearance of gender. Models of perceived gender are also important for studying how humans discriminate gender. As different expressive weights may be required for different recognition contexts (physical or perceived gender), the approach is designed to automatically learn the weight values for a specific context of labeled training data.

We demonstrate the applicability of our framework for gender recognition by training and evaluating our model with 40 point-light walkers (20 female, 20 male) labeled with their true physical gender and also with their perceived gender (obtained from a perceptual classification task). Results of our expressive PCA model show a higher classification rate attained for both contexts than with a standard non-expressive PCA model (with no weights).

Related work

Gender recognition from point-lights has received much attention during the past few decades. The first major experiment was presented in Kozlowski and Cutting (1977) with six walkers (three females, three males) of approximately the same height and weight recorded at a sagittal view. They demonstrated that human observers could classify the gender of the walkers with an average recognition rate of 63%, including one female walker who was identified as male by most participants. Alterations such as varying the arm swing, changing the walking speed, and occluding portions of the body were examined and found not to significantly influence recognition performance. Interestingly, they suggest that gender recognition is even possible from viewing only two moving ankle points (Kozlowski & Cutting, 1977; Kozlowski & Cutting, 1978).

A further study of gender recognition from point-lights was presented in Barclay et al. (1978), where temporal and spatial factors were examined. It was reported that successful gender recognition required exposure to approximately two walking cycles. The rendering speed was also found to have a strong influence over recognition. When the movements were recorded at normal walking speeds, but played back at an abnormally slow rate (about 3 times slower), the recognition rate dropped to almost chance level. Degraded displays, in which the point-lights were diffused into one bright blob, were also examined and shown to degrade recognition performance to chance level.

The effect of inversion on the point-lights was also investigated in Barclay et al. (1978). Interestingly, the gender assignments were significantly reversed (i.e., male walkers were seen as females, and vice versa). A view-based explanation was proposed based on the shoulder-hip ratio, in which men tend to have broader shoulders and smaller hips than women. With a perceptual bias to upright figures, the shoulder points in the inverted display would be perceived by the observer as hip points (and the inverted hip points as shoulder points), thus causing the interchange of the gender labels. Additional studies in Cutting et al. (1978) supported the shoulder-hip ratio concept and proposed a related center-of-moment feature of the torso.

The shoulder-hip ratio and center-of-moment features (Barclay et al., 1978; Cutting et al., 1978) are mainly based on the structural differences between male and female walkers. However, there are certainly dynamic visual features of movement that contribute to recognition. By setting structural and dynamic features into confliction using a synthetic point-light walker, experiments in Mather and Murdoch (1994) found that shoulder sway was an effective cue to gender at the frontal view (as supported by Murray, Kory, & Sepic, 1970). Structural and dynamic information were also compared in Troje (2002), where dynamic-only stimuli (movements applied to average postures) produced better results than with structural information (postures using averaged motions).

Most of the aforementioned experiments on recognizing the gender of walkers were conducted using a side-view presentation of the walkers to observers. Other experiments have been conducted to examine the effect of view angle on gender recognition performance. In Hirashima (1999), Mather and Murdoch (1994) and Troje (2002), it was found that a frontal-view presentation of the walkers consistently provided better gender recognition results than at a side view.

In Montepare and Zebrowitz-McArthur (1988), gender recognition using different age groups of the point-light walkers were examined. Child, adolescent, young adult, and older adult walkers showed an average recognition rate of 55%, 70%, 64%, and 69%, respectively. However, in a second experiment with a different group of adults, the average recognition rate dropped to 46%. Experiments in

Pollick, Lestou, Ryu, and Cho (2002) examined other point-light movements, such as knocking, lifting, and waving, which resulted in chance-level performance (a Neural Network demonstrated better performance). An average gender recognition rate of 75% was reported in Runeson and Frykholm (1983) for complex actions including walking, sitting, rising, standing on a chair, and jumping. They additionally explored the influence of the actor's natural, emphatic, and deceptive gender movement intention, which showed that the natural movements yielded the best gender recognition rate (86%). Similar actions were examined in Crawley, Good, Still, and Valenti (2000) using young children (4–5 years old), but resulted in unreliable gender recognition performance.

A recent pattern recognition approach for gender recognition was presented in Troje (2002), where a two-stage PCA framework was implemented to decompose male and female walking data into an Eigenspace, from which a linear classifier was used for gender recognition. The data consisted of three-dimensional (3D) motion-capture trajectories of 40 walkers (20 females, 20 males). The first PCA was applied to each walker to decompose the motion pattern into a posture basis. A walker description vector was constructed by concatenating the mean posture, first 4 posture components, and 4 sinusoidal parameters for modeling the periodic nature of the projection coefficients in the posture basis. A second PCA was then applied on the 40 description vectors (one for each walker) to further reduce the walker dimensionality. For gender recognition, a linear classifier was applied to the projection coefficients of the walkers in the second PCA space. The approach yielded a recognition rate of 92.5%.

Style analysis

In relation to gender recognition, much work in computer vision and computer animation/modeling has been devoted to the general modeling of movement styles.

In computer vision, a Parameterized-HMM was used by Wilson and Bobick (1999) to model spatial pointing gestures by adding a global variation parameter in the output probabilities of the HMM states. A bilinear model was used in Tenenbaum and Freeman (1997) for separating perceptual content and style parameters, and was demonstrated with extrapolation of fonts to unseen letters and translation of faces to novel illuminates. In Davis (2001), an approach to discriminate children from adults based on variations in relative stride length and stride frequency over various walking speeds was presented. Additionally, in Davis and Taylor (2002), the regularities in the walking motions for several people at different speeds were used to classify typical from atypical gaits. Morphable models were employed in Giese and Poggio (2000) to represent complex motion patterns by linear combinations of prototype sequences and used for movement analysis and synthesis. In Brand (2001), non-rigid objects were modeled from video by using singular value decomposition (SVD) with constraint and uncertainty factors. Analytical global transformations were employed in Yacoob and Black (1999) for recognizing atomic activities using PCA. A method for recognizing skill-level was presented in Yamamoto, Kondo, Yamagiwa, and Yamanaka (1998) to determine the ability of skiers by ranking properties such as synchronous and smooth motions.

In computer animation and modeling, a Fourier-based approach with basic and additional factors (walk; brisk) was employed in Unuma, Anjyo, and Takeuchi (1995) to generate human motion with different emotional properties (e.g., a happy walk). An HMM with entropy minimization was used by Brand and Hertzmann (2000) to generate different state-based animation styles. An N-mode factorization of motion-capture data for extracting person-specific motion signatures was described in Vasilescu (2001) to produce animations of people performing novel actions based on examples of other activities. SVD was used in Mason, Gomez, and Ebner (2001) to model reach-to-grasp hand postures. A movement exaggeration model using measurements of the observability and predictability of joint angle trajectories was presented in Davis and Kannappan (2002) to warp motions at one effort into increasing efforts using only selected trajectories. An approach using PCA to represent animation sequences was presented in Alexa and Muller (2000). In Chi, Costa, Zhao, and Badler (2000), the EMOTE character animation system used the effort and shape components of Laban Movement Analysis to describe a parameterization for generating natural synthetic gestures with different expressive qualities.

Prior work on expressive three-mode PCA

We outlined the fundamental three-mode PCA concepts using expressive features for style analysis in Davis et al. (Davis, Gao, & Kannappan, 2002; Davis & Gao, 2003a; Davis & Gao, 2003b). The method was examined with walking style variations caused by carrying load and pace. In Davis and Gao (2003a), we additionally conducted a limited initial experiment on the physical gender difference of walkers. In the present work, we present a special case of the general three-mode framework for recognizing the gender of point-light walkers. Specifically, we examine physical and perceptual recognition contexts for several point-light walkers. Also, we present new experiments related to perceptual parameterization of the walkers using perceptual classification experiments.

Our approach to gender recognition is most related to the PCA gender classification method of Troje (2002), yet there are important differences. First, we use 2D, rather than full 3D, trajectories for the experiments. Second, we normalize the walking speed and height of the walkers to remove these biases on the gender recognition task. Third, because our PCA model is based on a single three-mode factorization of the posture, time, and gender, our representation enables us to embed an expressive weight on each point-light trajectory within the PCA representation to adapt the estimation of the gender labels to the training

data (similar to a neural network approach). Thus our approach can be easily tuned to different recognition contexts (actual vs. perceived gender).

Theoretically, a similar weighting scheme could be used in other approaches. However, with a step-wise factorization (Troje, 2002), the walker data are typically rasterized and represented with basis vectors and projection coefficients. We feel that operating the expressive feature approach on these projection coefficients would be relevant only if the projections on the basis vectors capture some inherent, meaningful, decomposition of the original data (i.e., the decomposition provides meaningful representational units). Though this is possible, it is not generally clear that a PCA basis itself will provide a meaningful domain for feature weights when the data for each posture are rasterized together. Fourier analysis has a similar concern (though low and high frequency components could be important features, as shown in Unuma et al., 1995), and a frequency-based analysis may not be best-suited to non-periodic actions (e.g., lifting or throwing). In our expressive three-mode approach, we assign weights to each trajectory, not to each posture basis vector, and thus a more semantic and part-wise link to the actual motion is retained.

Expressive three-mode PCA model

Human movements can be visually described as specific body postures changing sequentially over time. Thus, in its most basic sense, movements have two visual modes: posture and time. If we further consider a stylistic component associated with the movement, such as the variation in walking due to gender, we have an additional third mode (one can easily argue for additional modes). In our approach, we exploit this tri-modal nature of female/male walkers (posture, time, and gender) with an efficient three-mode PCA representation that is suitable to incorporating tunable weights on trajectories to drive the recognition process to context-based matching criteria.

The data for multiple stylistic performances of a particular action can be naturally organized into a 3D cube $\mathbf{Z}$ (see Figure 1a), with the rows in each frontal plane/matrix $\mathbf{Z}_k$ composed of the point-light trajectories (segmented and normalized to a fixed length) for a particular style index k. The matrix data for each variation k could alternatively be rasterized into a column vector and placed into an ordinary two-mode matrix (each column a style example), but this simply ignores the underlying three-mode nature of the data (posture, time, and style).

Many times it is preferable to reduce the dimensionality of large data sets for ease of analysis (or recognition) by describing the data as linear combinations of a smaller number of latent, or hidden, prototypes. PCA and SVD are standard methods for achieving this data reduction, and have been successfully applied to several two-mode (matrix) problems (e.g., Alexa & Muller, 2000; Black, Yacoob, Jepson, & Fleet, 1997; Bobick & Davis, 1996; Li, Dettmer, & Shah, 1997; Murase & Nayar, 1995; Turk & Pentland, 1991; Yacoob & Black, 1999).

Three-mode factorization

Three-mode factorization (Tucker, 1966) is an extension of the traditional two-mode PCA/SVD in that it produces three orthonormal basis sets for a given three-mode data set arranged in a 3D cube $\mathbf{Z}$ (see Figure 1a). For gender recognition, we consider the style dimension as a binary gender mode of FEMALE or MALE. We begin by reducing the cube $\mathbf{Z}$ to a prototype data cube using two walking sequences, matrices $\bar{\mathbf{Z}}_f$ and $\bar{\mathbf{Z}}_m$, that represent the average female and average male walking styles. Each prototype is constructed by averaging multiple walking examples of each gender class. To ensure proper alignment when averaging, one walk cycle of each example is extracted (at the same walking phase), height-normalized, and time-normalized to a specific duration N. Further details of this preprocessing stage are presented later.

With a total of M point-light trajectories per person and N frames in the sequence (time-normalized walk cycle), each prototype is represented as a trajectory matrix of size $M \times N$. We then subtract the prototype mean $(\bar{\mathbf{Z}}_f + \bar{\mathbf{Z}}_m)/2$ from the two gender prototype matrices $\bar{\mathbf{Z}}_f$ and $\bar{\mathbf{Z}}_m$ and place them into the first and second (last) frontal plane of the cube $\bar{\mathbf{Z}}$ (see Figure 1b). The dimensionality of $\bar{\mathbf{Z}}$ is therefore $M \times N \times 2$.

The three-mode factorization of $\bar{\mathbf{Z}}$ decomposes it into three orthonormal matrices $\mathbf{P}$, $\mathbf{T}$, and $\mathbf{G}$ that span the column (posture), row (time), and slice (gender) dimensions of the cube (see Figure 1c). The core $\mathbf{C}$ is a matrix that represents the complex relationships of the components in $\mathbf{P}$, $\mathbf{T}$, and $\mathbf{G}$ for reconstructing $\bar{\mathbf{Z}}$. The desired column and row spaces can be found using SVD. We outline the technique in Appendix A.

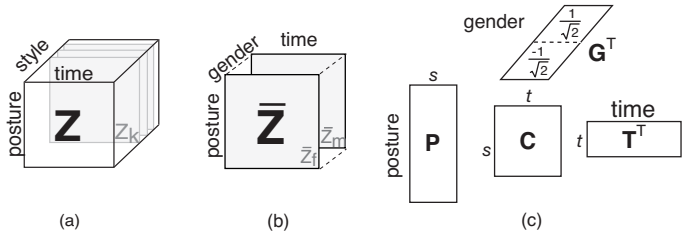

Figure 1. (a). General three-mode arrangement of stylistic motion data. (b). Three-mode arrangement of two gender prototypes. (c). Three-mode factorization of gender prototypes.

The posture basis $\mathbf{P}$ is able to represent any body posture (of point-lights) at any particular time for either gender prototype (i.e., column basis for $\bar{\mathbf{Z}}$). The time basis $\mathbf{T}$ represents any temporal trajectory (of any point-light) for either gender prototype (i.e., row basis for $\bar{\mathbf{Z}}$). Lastly, the gender basis $\mathbf{G}$ represents the gender-related changes be-

tween the two prototypes for any posture at any particular time (i.e., slice-line basis for $\bar{\mathbf{Z}}$).

Typically, each mode needs only to retain its first few components (meeting some modal variance criteria) to capture most of the fit to $\bar{\mathbf{Z}}$. As there are only two mean-subtracted gender prototypes, the normalized gender basis is constrained to be $\mathbf{G} = [-1, 1]^{\mathrm{T}} / \sqrt{2}$, signifying the female ($-1/\sqrt{2}$) and male ($1/\sqrt{2}$) prototype sides of the mean.

The complete three-mode factorization of $\bar{\mathbf{Z}}$, in flattened matrix form, can be concisely written as

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

where $\otimes$ is the Kronecker product (Kroonenberg, 1983) and $[\bar{\mathbf{Z}}_f \mid \bar{\mathbf{Z}}_m]$ is a matrix with the columns of $\bar{\mathbf{Z}}_f$ followed by the columns of $\bar{\mathbf{Z}}_m$. The core matrix $\mathbf{C}$ can then be solved by simply re-arranging Equation 1 as

$$\mathbf{C} = \mathbf{P}^{\mathrm{T}}[\bar{\mathbf{Z}}_f \mid \bar{\mathbf{Z}}_m](\mathbf{G}^{\mathrm{T}} \otimes \mathbf{T}^{\mathrm{T}})^{\mathrm{T}}, \quad (2)$$

where $\mathbf{C}$ need not be diagonal, as is required in two-mode PCA/SVD. Related methods for solving this three-mode factorization can be found in Kroonenberg and Leeuw (1980) and Vasilescu and Terzopoulos (2002).

Three-mode gender estimation

From Equation 1, each gender prototype ($\bar{\mathbf{Z}}_f, \bar{\mathbf{Z}}_m$) can be reconstructed as

$$\bar{\mathbf{Z}}_{\{f,m\}} = \mathbf{PC}g_{\{f,m\}}\mathbf{T}^{\mathrm{T}}, \quad (3)$$

where $\bar{\mathbf{Z}}_{\{f,m\}}$ corresponds to either gender prototype. The gender parameter g signifies the gender with $g_f = -1/\sqrt{2}$ for FEMALE and $g_m = 1/\sqrt{2}$ for MALE

$$\bar{\mathbf{Z}}_f = \mathbf{PC}\frac{-1}{\sqrt{2}}\mathbf{T}^{\mathrm{T}} \quad (4)$$

$$\bar{\mathbf{Z}}_m = \mathbf{PC}\frac{1}{\sqrt{2}}\mathbf{T}^{\mathrm{T}}. \quad (5)$$

We can write Equation 3 as a summation of three-mode basis elements and isolate the gender parameter from the remaining factored terms with

$$\bar{\mathbf{Z}}_{ij\{f,m\}} = \sum_{p=1}^{s}\sum_{q=1}^{t}\mathbf{P}_{ip}\mathbf{C}_{pq}g_{\{f,m\}}\mathbf{T}_{jq} \quad (6)$$

$$= g_{\{f,m\}}\left(\sum_{p=1}^{s}\sum_{q=1}^{t}\mathbf{P}_{ip}\mathbf{C}_{pq}\mathbf{T}_{jq}\right) \quad (7)$$

$$= g_{\{f,m\}} \cdot \alpha_{ij}, \quad (8)$$

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

To determine the gender for a new walker within this framework, we need only to estimate its corresponding gender parameter. For this, a minimization of the three-mode PCA reconstruction error for the new walker can be employed. Following Equation 8, the unknown gender parameter $\hat{g}$ for a new walker $\mathbf{Y}$ (already mean-subtracted with the model) can be estimated by finding the value of $\hat{g}$ that minimizes the sum-of-squared-error (SSE) reconstruction

$$\mathcal{F} = \sum_{i=1}^{M}\sum_{j=1}^{N}(\mathbf{Y}_{ij} - \hat{g} \cdot \alpha_{ij})^2. \quad (9)$$

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

$$\hat{g} = \frac{\sum_i \sum_j \mathbf{Y}_{ij} \cdot \alpha_{ij}}{\sum_i \sum_j \alpha_{ij}^2}, \quad (10)$$

where the gender parameter is computed by the normalized projection of $\mathbf{Y}$ onto the basis. The final gender 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 gender prototype).

Expressive three-mode gender estimation

Gender estimation using Equation 10 could have equally been achieved by rasterizing the gender prototype data into a $MN \times 2$ matrix (each column is a rasterized gender prototype), performing a standard two-mode PCA, and estimating the gender parameter for a new walker by computing and thresholding its projection coefficient. The three-mode formulation (Equation 10), however, enables us to easily embed tunable weight factors (on trajectories) to influence the estimation of the gender parameter.

The minimization of Equation 9 seeks a value of the gender parameter $\hat{g}$ that reduces the reconstruction error in a squared-error manner. Hence, any point-light trajectories having large magnitude differences from the model will significantly influence the minimization process (outliers are a common problem in SSE minimizations). However, only certain trajectories may carry the most expressive and consistent information regarding the gender differences across several walking examples. Furthermore, the most expressive trajectories could have smaller magnitude differences in comparison to the remaining trajectories, thus attenuating their impact in an SSE gender estimation process. What is needed is a method to weight trajectories differently to enable the most expressive trajectories to drive the estimation process.

Using the three-mode reconstruction error equation (Equation 9), we introduce a weight factor in the range 0–1 on each of the M point-light trajectories with

$$\mathcal{F} = \sum_{i=1}^{M}\mathcal{E}_i\sum_{j=1}^{N}(\mathbf{Y}_{ij} - \hat{g} \cdot \alpha_{ij})^2. \quad (11)$$

The new expressive gender parameter estimation is given by

$$\hat{g} = \frac{\sum_i \mathcal{E}_i \sum_j \mathbf{Y}_{ij} \cdot \alpha_{ij}}{\sum_i \mathcal{E}_i \sum_j \alpha_{ij}^2} \quad (12)$$

$$= \frac{\sum_i \mathcal{E}_i \Delta_i}{\sum_i \mathcal{E}_i \sum_j \alpha_{ij}^2} \quad (13)$$

$$= \sum_i \tilde{\mathcal{E}}_i \Delta_i \,, \quad (14)$$

where $\Delta_i = \sum_{j=1}^{N} \mathbf{Y}_{ij} \cdot \alpha_{ij}$. As the denominator $\sum_i \mathcal{E}_i \sum_j \alpha_{ij}^2$ in Equation 13 is a constant for a given set of factors $\mathcal{E}_i$, we fold this term into the final "expressive weights" $\tilde{\mathcal{E}}_i$ in Equation 14. If we set each expressive weight to $1/(\sum_i \sum_j \alpha_{ij}^2)$ in Equation 14, the resulting gender parameter estimation reverts to the previous SSE method (Equation 10). However, with non-uniform values for $\tilde{\mathcal{E}}_i$, the approach is capable of producing other non-SSE gender estimations according to a specific recognition context.

Learning expressive weights

We present a learning-based method to determine the appropriate values of the expressive weights $\tilde{\mathcal{E}}_i$ by minimizing a second error function that compares the computed gender parameters $\hat{g}$ (using Equation 14) with labels ($\bar{g} = \pm 1/\sqrt{2}$) pre-assigned to K different training examples

$$\mathcal{J} = \sum_{k=1}^{K} (\bar{g}_k - \hat{g}_k)^2 \quad (15)$$

$$= \sum_{k=1}^{K} (\bar{g}_k - \sum_{i=1}^{M} \tilde{\mathcal{E}}_i \cdot \Delta_{ik})^2 \,. \quad (16)$$

To solve for the expressive weights in Equation 16, we employ a fast iterative gradient descent algorithm (Burden & Faires, 1993) of the form

$$\tilde{\mathcal{E}}_i(n+1) = \tilde{\mathcal{E}}_i(n) - \eta(n) \cdot \frac{\partial \mathcal{J}}{\partial \tilde{\mathcal{E}}_i} \,, \quad (17)$$

with the gradients $\partial \mathcal{J} / \partial \tilde{\mathcal{E}}_i$ computed over the K training examples

$$\frac{\partial \mathcal{J}}{\partial \tilde{\mathcal{E}}_i} = -2 \sum_{k=1}^{K} \Delta_{ik} (\bar{g}_k - \sum_{m=1}^{M} \tilde{\mathcal{E}}_m \cdot \Delta_{mk}) \,. \quad (18)$$

The learning rate η is re-computed at each iteration (via interpolation of the error function) (Burden & Faires, 1993) to yield the best incremental update.

The expressive weights are initialized to the default SSE formulation, where each weight is initially set to $1/(\sum_m \sum_j \alpha_{mj}^2)$, which therefore guarantees a final solution (even at a local minima) with a smaller error than produced with SSE (i.e., with no expressive weights). The weights are also confined to be positive (by definition) in each iteration. Experiments conducted with random values in the locus of the SSE values ($\mathcal{E}_i$ randomly set between 0 and 1, and the expressive weights $\tilde{\mathcal{E}}_i$ initialized to $\mathcal{E}_i / (\sum_m \mathcal{E}_m \sum_j \alpha_{mj}^2)$ showed fairly consistent convergence in experiments with our data. Following termination of Equation 17, the gender parameter for a walker can be estimated with Equation 14 using the newly learned expressive weights.

The general gradient descent algorithm determines a local minimum for a multi-parameter error function by searching through the parameter space to find values that yield the minimum error. The algorithm evaluates the error function with the current parameter estimates and updates the parameters by a small amount in the opposite direction of the error gradient to reduce the error function. This updating process is repeated until it converges or reaches a maximum number of iterations. It is certainly possible to use other minimization techniques to estimate the expressive weights in our formulation. If the training set contains enough examples (more than the number of expressive weights), the error function could in fact be solved linearly. However, the matrix inversion step could produce negative weights.

Interpretation of expressive weights

Our framework offers a method to learn numeric weightings of representational units to bias the estimation of the gender parameter from a given set of labeled training data. In essence, the approach combines PCA and Neural Network learning techniques into a single framework.

In the experiments presented in this work, we employ low-level 2D position trajectories as the representational units. Those trajectories assigned a zero-valued expressive weight can be interpreted as non-expressive gender features in the model. For the remaining trajectories with non-zero expressive weights, we cannot definitively state that trajectories with larger weights are given more influence in the classification task. The magnitude differences between the input and reconstructed trajectories directly influence the magnitude of the weights during the numerical minimization procedure. One possibility to correct this might be to first normalize the trajectory data in some manner as to remove any effects due to scale across the different trajectories.

Therefore, with the current trajectory representation, we make no particular claim that the resulting non-zero expressive weight magnitudes are indicative of what is being used in human perception. However, if a suitable perceptually-based representation, perhaps using normalized joint-angles, is employed in the framework, the resulting weight magnitudes computed with the approach may potentially show high-level correlations indicative to what humans are using for discrimination. This will require further investigation.

Walking data

To conduct the experiments, we employed a set of 20 female and 20 male motion-capture walking movements collected by N. Troje at the BioMotionLab of Ruhr-University, Bochum, Germany. The participants were mostly students and staff in the Psychology Department, aged between 20 and 38 years (average age of 26 years). A set of 38 retroreflective markers was attached to the body of each walker using a standard marker configuration for human figure motion-capture. Participants were each asked to walk on a treadmill and adjust the speed to the most comfortable setting. Ten gait cycles were recorded after the participant was walking for at least 5 min (the participant was not notified when recording was to begin). A Vicon motion-capture system with 9 high-speed CCD cameras was used to capture the 3D marker positions within 1 mm spatial resolution at 120-FPS temporal sampling.

This motion-capture data was previously used (see Troje, 2002), however the data for each person were transformed (using BodyBuilder biomechanical modeling software) into a stick-figure skeleton consisting of 15 virtual 3D points located approximately at the major joint locations of the body. To eliminate any artifacts that may arise when computing such skeletons, we instead chose to select most of our point-lights directly from the original marker set. We first selected 10 markers at the major arm and leg joint locations. Then we averaged the 4 head markers (left-front, right-front, left-rear, right-rear) into a single head point. Similarly, the left and right hip markers (left-front, left-rear; right-front, right-rear) were averaged on each side (producing left-hip and right-hip points). The result was a set of 13 point-lights mostly located on the body surface (except for the head and hips).

To suppress any noise from the motion-capture system, we smoothed the trajectories with a fifth-order, zero-phase forward-and-reverse lowpass Butterworth filter with cut-off at 6 Hz. Each walking sequence was then rotated to have all participants facing the same forward direction (positive Z-axis). The center-of-rotation (root location) for each person was selected as the average center position between the hips throughout the walking sequence. The angle of rotation was computed such that the average root orientation was facing directly forward.

To remove any potential bias due to person height, the stature of each person was normalized by scaling the point-lights with the average length of the person's left and right tibias (distance between knee and ankle markers). Additionally, one walk cycle was extracted from each sequence (at the same walking phase) by detecting cyclic curvature peaks in the left-knee trajectory. Each walk cycle was then time-normalized to a fixed duration using spline interpolation to remove any bias of walking speed on the gender recognition task. Each cycle was time-normalized to $N = 50$ frames to avoid under-sampling (longest cycle sequence of the walkers was 44 frames at 30 FPS).

To make a continuously repetitive walk cycle (for the perceptual experiments), we used a simple approach that removes the discontinuity between the last and first frame of the cycle for each trajectory $x(t)$ by distributing the error $\delta = x(1) - x(N)$ throughout the trajectory as

$$\tilde{x}(t) = x(t) + \frac{(t-1)\delta}{N}, \qquad (19)$$

where t denotes the frame number ($1 \le t \le N$). The approach distributes the discontinuity using small shifts throughout the trajectory to align the starting and ending positions without the loss of high-frequency information. This simple, yet effective, method produces seamless walking cycles without noticeable visual distortion (other Fourier-based techniques could also be employed).

Lastly, the 3D trajectories were orthographically projected into 2D at the frontal view. The forward viewing direction was previously found to yield the best gender discrimination rate by human observers (Hirashima, 1999; Troje, 2002). The frontal view also avoids the problem of estimating the shoulder-hip ratio or center-of-moment (Barclay et al., 1978; Cutting et al., 1978). Example point-light images from our dataset for three female and three male walkers are shown in Figure 2.

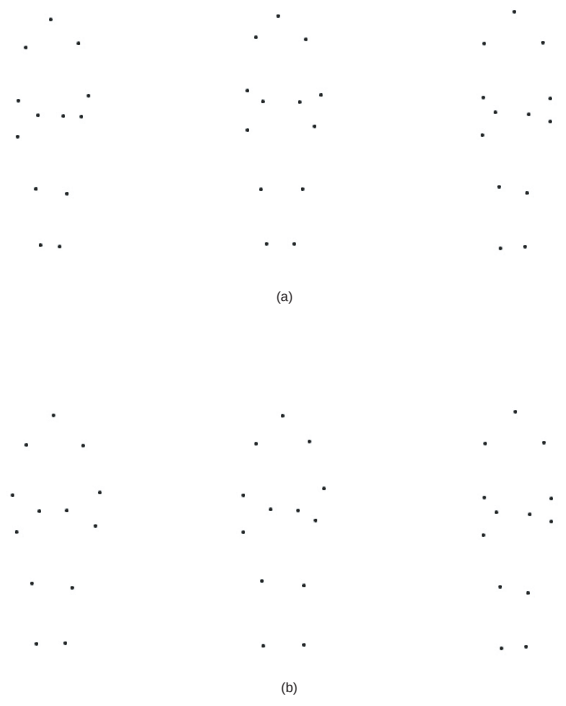

Figure 2. Point-light images at the frontal view. (a). Three female walkers (#2, #8, and #16). (b). Three male walkers (#22, #32, and #37).

Gender recognition by human observers

Before testing our three-mode approach for gender recognition, we first examined the capability of human observers to recognize the gender of our point-light walkers. The perceptual classification labels produced from this experiment will also be used to train our expressive three-mode model in the perceptual recognition context.

Participants

Fifteen students (5 female, 10 male) from Ohio State University were recruited as participants for the experiment. All English-speaking participants had normal or corrected-to-normal vision. Their ages ranged from 21 to 34 years (average age of 25 years). Some participants had previously been exposed to point-light stimuli, but not to the displays used in this experiment.

Methods

A computer program was implemented to display the 40 female/male walkers and to collect the observer responses of the perceived gender of the walkers. Each sequence was rendered as black points against an off-white background (see Figure 3).

Because the time-normalized walk cycles (length of $N = 50$ frames) appear abnormally slow if rendered at 30 FPS, we used a slightly faster rendering speed of 36 FPS determined from the longest natural cycle time of the walkers (1.4 s). Each sequence was looped continuously while presented to the observer. The height of each point-light walker was scaled to 70% of the screen resolution height (1280 × 1024 resolution, with 20-in. diagonal viewable monitor). The root location of each walker was randomly positioned within a small circle at the center of screen (with radius 10% of the screen resolution height). These display parameters were used to prevent any explicit position or size comparison between the walkers. The point-light display was generated using C++ and OpenGL with anti-aliasing. Each observer was seated approximately 60 cm from the monitor, which corresponded to a visual angle of approximately 20 deg for the height of the point-light figure.

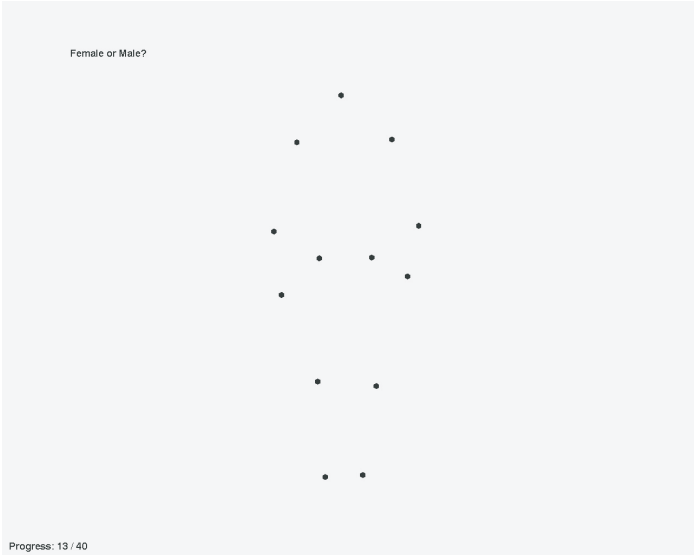

Figure 3. Screen-shot of the computer display used for the gender-recognition task.

For each displayed walker sequence, the participant was asked to select a gender label using the keyboard, pressing the 'F' key to select FEMALE or 'M' for MALE. To confirm/save the choice and load the next stimuli, the participant was required to press the 'Enter' key. The 40 sequences were presented in random order for each trial. The progress was shown in the bottom-left of the computer display (though not required for the experiment), and no time restriction was enforced. Each walker labeled by the participant as FEMALE was assigned a numeric label of −1, and each sequence labeled as MALE was assigned +1. Each participant was paid $5, and an additional incentive of 50 cents per correct gender selected after exceeding random-chance performance was paid at the end of the experiment (maximum possible payment of $15).

To ensure that each participant was able to perceive the moving point-lights as a walking person (required before determining the gender), we introduced a preview stage before the actual experiment. A sample point-light walking sequence was shown to each participant and told that it contains a person walking on a treadmill with markers attached to the major limbs. The participant must verbally confirm the presence of a walking person before beginning the experiment (all participants could easily recognize the display). As not to bias the gender recognition task, walker #21 (male) was selected for the preview, because it resulted in an ambiguous gender assignment from an earlier pilot study with 7 observers.

Results

The average recognition rate for the 40 walking sequences across all 15 observers was 69%. The result is significantly above chance performance ($t(39) = 5.52$, $p < .001$). Previous experiments employing a frontal view of walkers (as in this experiment) reported rates of 64% (Hirashima, 1999) and 76% (Troje, 2002).

Though the experiments in Troje (2002) were conducted using the same walkers as in this experiment (both datasets were derived from the same motion-capture data), the difference in recognition rates could potentially be explained by different stimuli presented to the observers. In Troje (2002), biomechanical modeling software was employed to create a virtual stick-figure skeleton from the full marker set. We used 13 point-light trajectories, of which 10 point-lights were directly chosen from the original marker data (the remaining 3 were averaged from other original

markers). Additionally, Troje (2002) presented several walking cycles, whereas we only presented one (looped) walk cycle at a fixed speed.

To examine the individual walker results, we computed a gender consistency value for each walker by averaging the numerical values (±1) assigned by the 15 observers. 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 gender consistencies from the perceptual experiment with the 40 point-light walkers in Figure 4. Walkers #1-20 are true females, and walkers #21-40 are true males. There appears to be a slight bias toward perceived maleness in the walkers. There were three walkers (#1, #4, #34) whose gender labels were unanimously selected by all the observers. Interestingly, walker #4 was a female that was labeled as male by all participants. This clearly demonstrates the potential differences between perceived and actual gender. Several other walkers were still difficult to label (with consistency values near zero).

Previous studies (Barclay et al., 1978; Cutting et al., 1978) have suggested the shoulder-hip ratio as a factor influencing gender recognition. We computed this ratio using the 2D width of the shoulder s and hip h in the first image of each walker sequence. We note that there could, however, be more discriminative structural information in later frames (though it should not change drastically at the frontal view). The average shoulder-hip ratio s/h was 1.71 ±.26 for females and 1.92 ±.14 for males, and were significantly different (two-tailed t test: $t(38) = 3.14$, $p < .01$). The differences in our shoulder-hip ratios with previous measurements (Cutting et al., 1978) are likely due to the inward placement of the front and back hip markers on the body (not at the maximal hip width). The two hip point-lights (left, right) were created by averaging the back and front hip markers on each side. Therefore the calculated hip width would be shorter than the actual hip width (thus increasing the shoulder-hip ratio).

The lack of a strong correlation between the shoulder-hip ratios for the walkers and the gender consistency values ($r = .34$) suggests that this factor alone does not account for the perceptual gender choices. We also computed the center-of-moment for the walkers, using $C_m = s/(s+h)$. Even though significantly different for females and males (two-tailed t test: $t(38) = 3.04$, $p < .01$), its correlation with the gender consistency values was also low ($r = .34$).

Comparison to static display

In addition to the dynamic point-light stimuli, we presented observers with only one frame from each point-light sequence to examine the influence of motion for the gender recognition task. We recruited 15 new students (3 female, 12 male) not involved with the previous experiment as participants. Their ages ranged from 20 to 29 years (average age of 24 years).

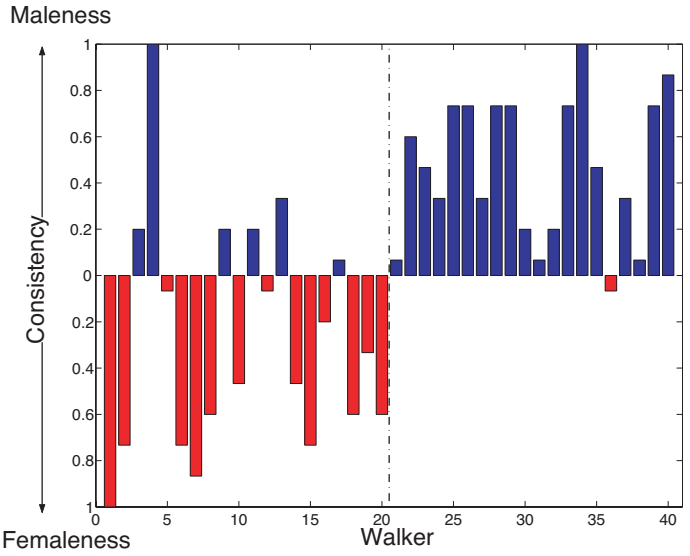

Figure 4. Gender consistency values for the 40 walking sequences (females: #1-20; males: #21-40).

A total of 40 single-frame point-light images of the female and male walkers comprised the stimuli. As the walking poses look very similar at the frontal view (as opposed to the sideview), we chose to employ only the initial frame in the walk cycle rather than multiple static frames (as in Kozlowski & Cutting, 1977). We again admit that static differences could potentially arise at other frames/poses. Examples images are shown in Figure 2. The same computer program (now only displaying a single frame), preview, and compensation method were employed as in the previous dynamic experiment. The gender selections were collected from the participants and averaged into gender consistency values (see Figure 5a).

As expected, few walkers were strongly identified with their true gender. The average recognition rate for the 40 static images across all 15 observers was 57%, and was above chance ($t(39) = 2.62$, $p < .05$). Many of the walkers were ambiguous to label given one static frame, yet walkers #2, #32, and #37 were correctly recognized at 87%, 87%, and 93%, respectively.

In comparison, the overall static and dynamic rates were significantly different (two-tailed t test: $t(78) = 2.62$, $p < .05$), with the static recognition rate almost 10% lower than achieved with the dynamic displays. We also calculated the absolute value of the difference between the static and dynamic consistency values (see Figure 5b). A large difference magnitude close to 2 for a walker indicates a strong gender inconsistency between the static and dynamic cases, and a value close to 0 indicates that the two stimuli provided a similar (strong or weak) gender consistency. Interestingly, several walkers (e.g., #7, #10, #15, and #29) had fairly strong gender conflictions between the static and dynamic cases. Overall, the dynamic stimuli appear to give more gender-related information than the single frame case.

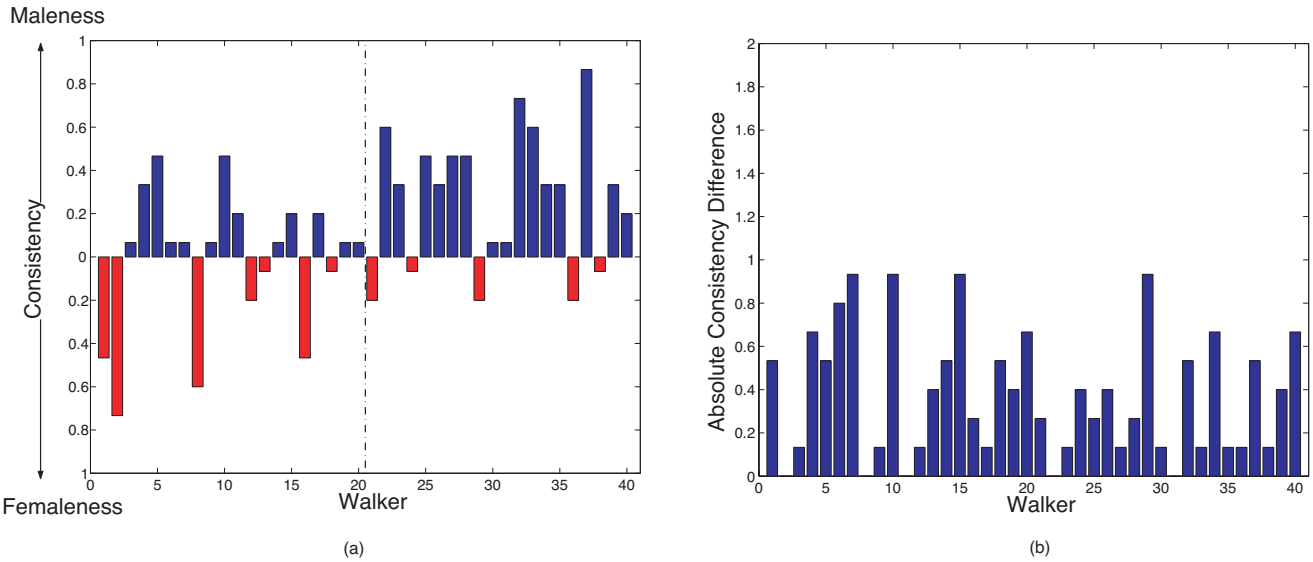

Figure 5. Static display experimental results. (a). Gender consistency values for the 40 static walking images. (b). Differences between static and dynamic consistency values.

Gender recognition using expressive three-mode PCA

We now examine our expressive three-mode PCA model for gender recognition with the same 40 female/male walker data used in the perceptual experiments. We demonstrate the learning flexibility of the framework by recognizing gender based on the actual physical label (*is* female/male) and the perceived label (*appears* female/male).

Recognizing physical gender

To compute the optimal three-mode PCA model and to avoid overfitting, we employed a leave-one-out cross-validation technique. In this method, we simultaneously varied the percentage modal fit of the three-mode **P** and **T** basis sets (posture and time) from 50% – 95% (in 5% increments). For example, an "85% modal fit" means that we accumulate the top basis vectors in the posture basis **P** until 85% of the variance in the data is captured. We apply the same criterion for the basis **T**. The gender basis **G** remains fixed to $[-1, 1]^T / \sqrt{2}$.

For each percent modal fit of **P** and **T**, we constructed 40 different models, each using 39 training examples by leaving one different example out of the set. For each model (39 examples at a particular modal fit), we created the gender prototypes, computed the three-mode PCA for the prototypes, and ran the learning algorithm to compute the expressive weights (examples labeled with their true gender). We empirically selected a limit of 1,500 iterations for the gradient descent learning algorithm as it provided satisfactory convergence of the expressive weights for our data set (in both recognition contexts). The training error for the model was computed by examining the sign of the computed gender parameter value for each of its 39 labeled training examples (–: FEMALE, +: MALE). The testing (validation) error for the model was similarly computed using only the single left-out example.

For each modal fit, we then computed the average training and testing errors of the 40 leave-one-out models. The cross-validation training and testing errors at each modal fit are shown in Figure 6a. To select the optimal modal fit for the data, we chose the fit (75%) that corresponded to the smallest average testing error (25%).

We then constructed a single expressive three-mode model at this modal fit. First, the prototypes were created from the full training set. Next, the basis sets **P** and **T** were computed at the selected modal fit (75%), and were of dimension 26 × 3 and 50 × 3, respectively (the core **C** was therefore of size 3 × 3). The resulting three-mode PCA captured 98% of the overall data variance in the two gender prototypes. The expressive weights for this model were generated by averaging the 40 sets of expressive weights computed at the selected cross-validation modal fit (75%). We show the average expressive weights ±1 *SD* in Figure 7. Some weights were zero, signifying that they were not relevant to the gender assignments. The larger magnitude weights appear to have a significant deviation across the 40 leave-one-out sets, showing the impact of the singular left-out examples. However, as previously mentioned, it is difficult to assign any high-level interpretation to the larger magnitude weights. The mapping of the 26 weights to the point-lights is shown in Figure 8.

Results

To evaluate the resulting expressive three-mode PCA model, we first computed the raw (unthresholded) gender parameters using Equation 14 for all 40 walkers, and com-

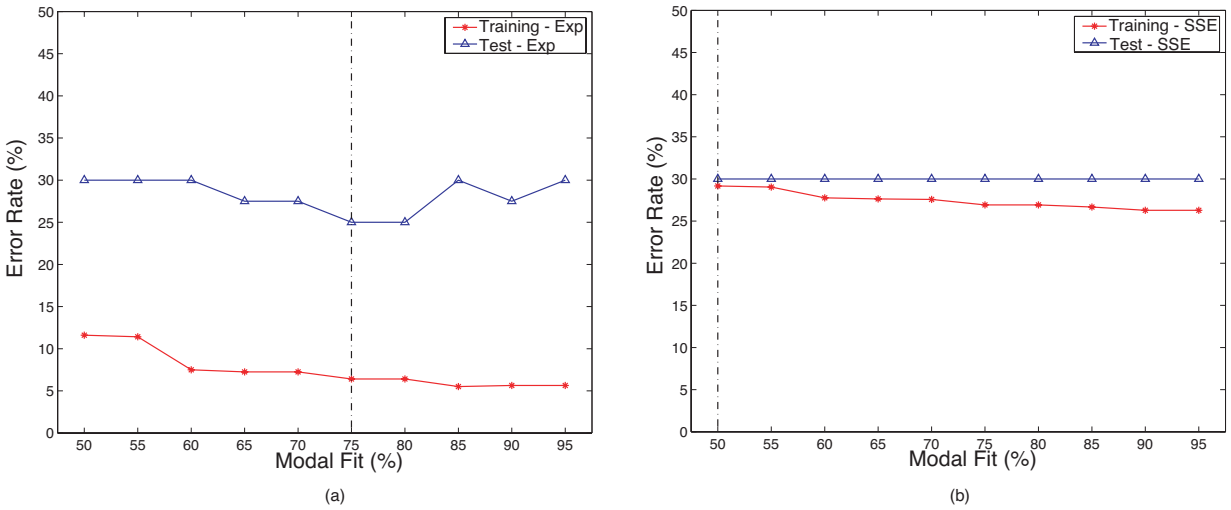

Figure 6. Cross-validation training and testing errors for different modal fits of physical gender. (a). Expressive model. (b). Non-expressive SSE model.

pared the results to the assigned $\pm 1/\sqrt{2}$ physical gender values. The target (training) labels and computed values from the expressive model are shown in Figure 9a.

For comparison, we employed the same cross-validation technique on a three-mode PCA model without any expressive weights (i.e., using the default SSE estimation). The cross-validation errors are shown in Figure 6b. A constant testing error is present, though different examples resulted in the errors across the modal fits. In this case, it appears that early generalization (at 50% fit) was achieved for the SSE model (with a 30% testing error). The unthresholded gender parameters produced from the final non-expressive SSE model and the target values are shown in Figure 9b.

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). Thresholding the gender parameter values at zero produced a 92.5% classification rate with our expressive model, and a much lower 70% classification rate for the non-expressive SSE version. Although the testing errors during cross-validation for the expressive and SSE models were similar (expressive model was slightly better), the training errors for the expressive model were significantly less than with SSE. Furthermore, we applied a *generalized* set of expressive weights (averaged from the 40 cross-validation models at the selected modal fit). Both of these

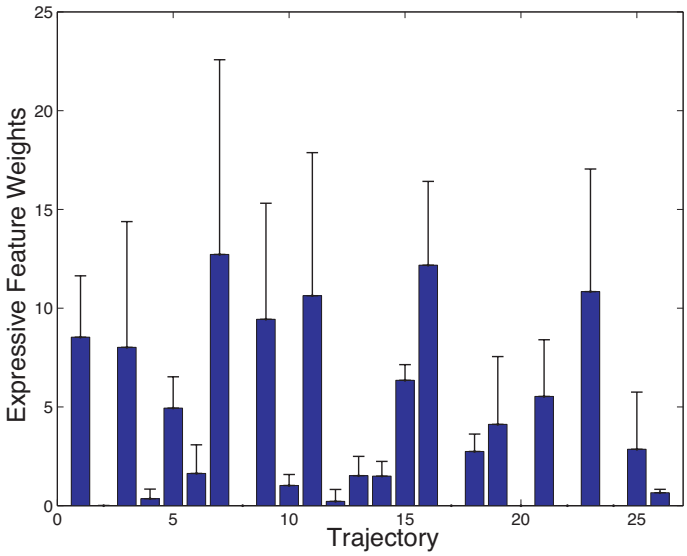

Figure 7. Average expressive weights ±1 *SD* from the cross-validation set for physical gender.

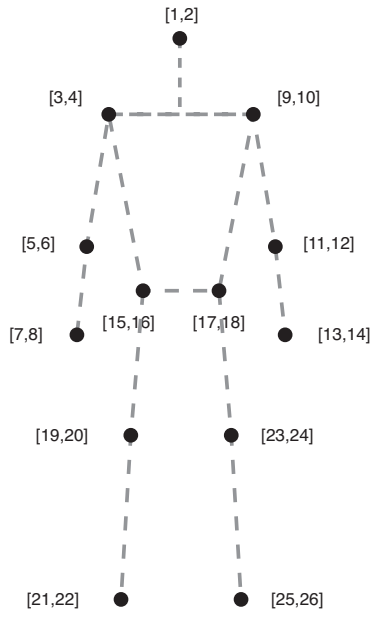

Figure 8. Point-lights [x,y] labeled with expressive weight index.

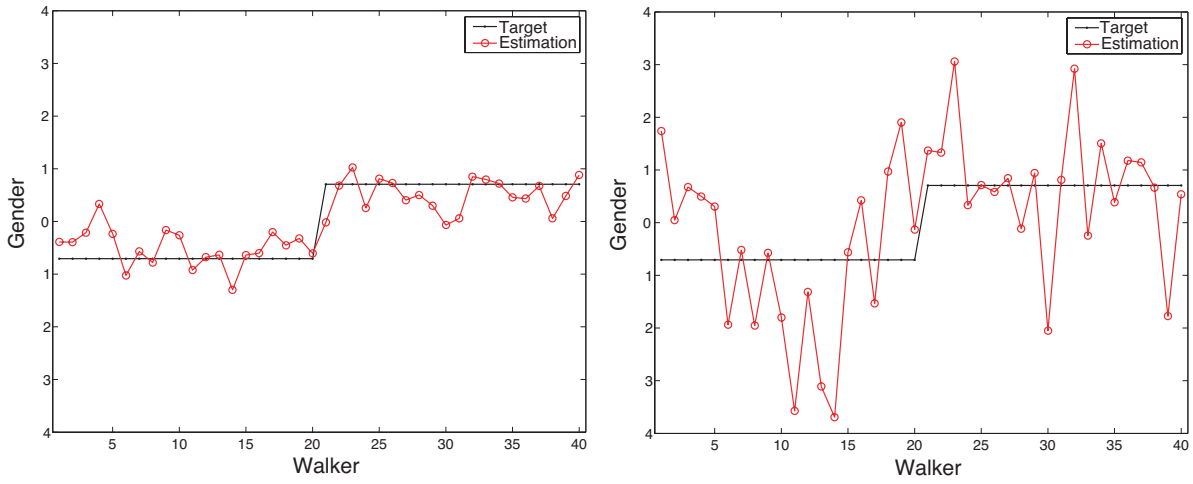

Figure 9. Physical gender parameter estimation results. (a). Expressive estimation. (b). Non-expressive SSE estimation. (walkers 1-20: female; 21-40: male)

factors enabled the expressive model to achieve a smaller recognition error than with SSE.

Recognizing perceptual gender

An advantage to our framework is that the model can adapt to a different labeling of the same underlying training data. To demonstrate this capability, we also trained our model to produce gender estimations more similar to the classification results attained from human observers of the data. We employed the gender consistency values from the earlier dynamic perceptual experiment to label the walking data. For each walker in the training set, we assigned a perceptual gender label of $-1/\sqrt{2}$ (FEMALE) if it had a negative gender consistency or $1/\sqrt{2}$ (MALE) otherwise. The perceptual training set resulted in 15 perceived-females (including 14 true females) and 25 perceived-males (including 19 true males).

We then used the same cross-validation technique (over multiple modal fits) outlined in the previous section to construct the optimal expressive PCA model. In this perceptual context, the two prototypes were constructed using the perceived-female and perceived-male classes (not the true gender). The average cross-validation training and testing errors for the expressive model at different modal fits are shown in Figure 10a. The optimal cross-validation set was found at an 80% modal fit (testing error of 28%).

For the optimal three-mode model, the prototypes were computed from all of the perceived-females and perceived-males. The basis sets $\mathbf{P}$ and $\mathbf{T}$ (at 80% modal fit) were of dimension 26×4 and 50×4, respectively (the core $\mathbf{C}$ was therefore of size 4×4). The resulting three-mode PCA captured 98% of the overall data variance in the two gender prototypes. The expressive weights were generated by averaging the 40 sets of expressive weights computed at the 80% cross-validation fit. We show the average weights ±1

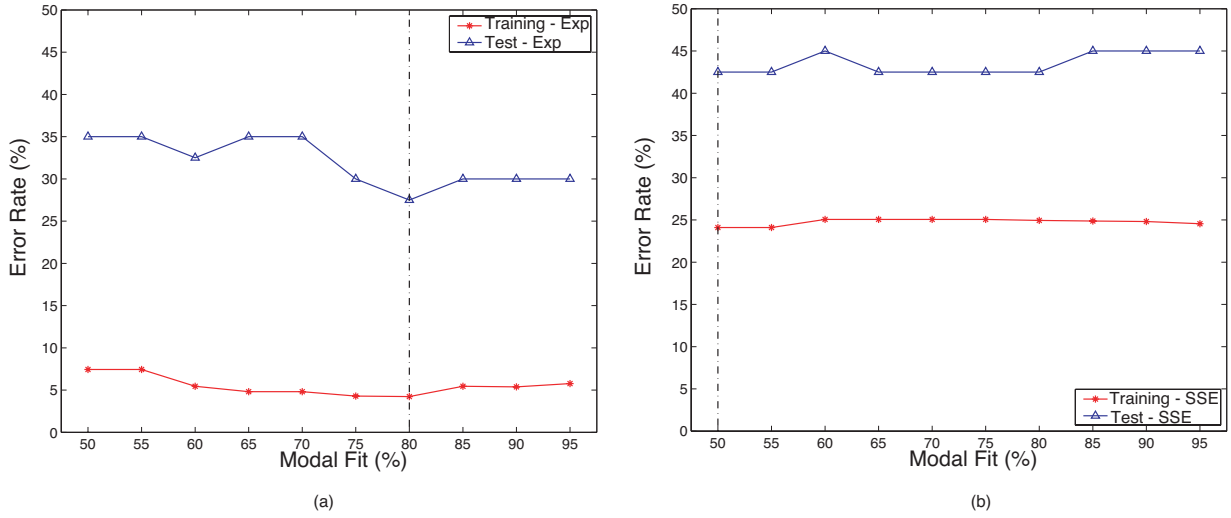

Figure 10. Cross-validation training and testing errors for different modal fits of perceptual gender. (a). Expressive model. (b). Non-expressive SSE model.

SD in Figure 11. As before, some weights are zero, and we also see a larger variation in the higher magnitude weights. As the training data for the physical and perceptual labels are in fact different, we expect the resulting weights to also be different. There is an unexplained asymmetry in the arms, though this may be mostly due to the variation in the cross-validation weights. However, we do not yet have a high-level interpretation of the cause for the weight differences between the two contexts.

Results

The resulting gender parameter estimations for the expressive model are shown in Figure 12a. For comparison, we also computed the optimal SSE model for the perceptual data using the cross-validation technique (see Figure 10b), and show its gender parameter estimations in Figure 12b. As in the previous physical gender case, our expressive model produced gender parameter values much closer to the desired perceptual values (average difference = .33, similar to the physical gender results) than did the alternative SSE estimation approach (average difference = .68). Thresholding the gender parameter values at zero produced a 90% classification rate for our expressive model and 77.5% for the non-expressive SSE model. The correlation of the expressive perceptual gender parameter with the gender consistency values was r = .89 (SSE correlation was r =.69).

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 and give the remaining walkers with high consistency magnitudes more emphasis when learning the expressive weights. Given the set of K training examples and their assigned perceptual genders $\bar{g}_k$, we slightly alter the previous matching error function (Equation 16) by using their consistency magnitudes ω_k to bias the minimization procedure to those examples having more reliable matches across the observers

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

The corresponding perceptual gradient is then

$$\frac{\partial \mathcal{J}_p}{\partial \tilde{\mathcal{E}}_i} = -2 \sum_{k=1}^{K} \omega_k \Delta_{ik} (\bar{g}_k - \sum_{m=1}^{M} \tilde{\mathcal{E}}_m \cdot \Delta_{mk}) . \quad (21)$$

This new gradient is used as before in the gradient descent procedure (Equation 17) 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 ex-

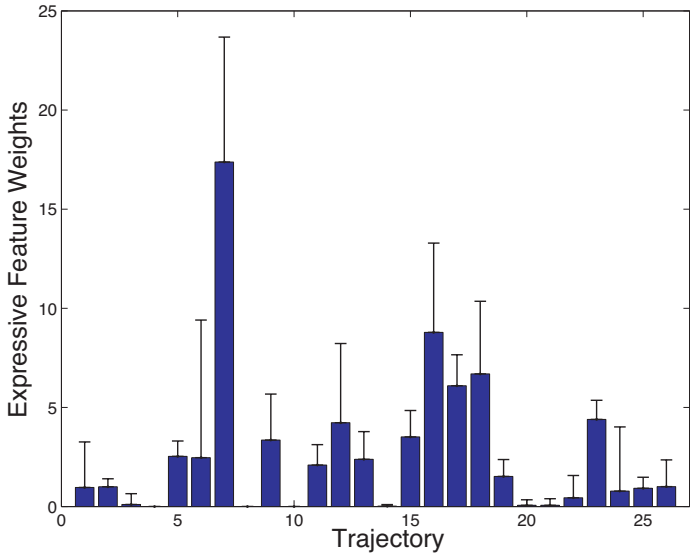

Figure 11. Average expressive weights ±1 *SD* from the cross-validation set for perceptual gender.

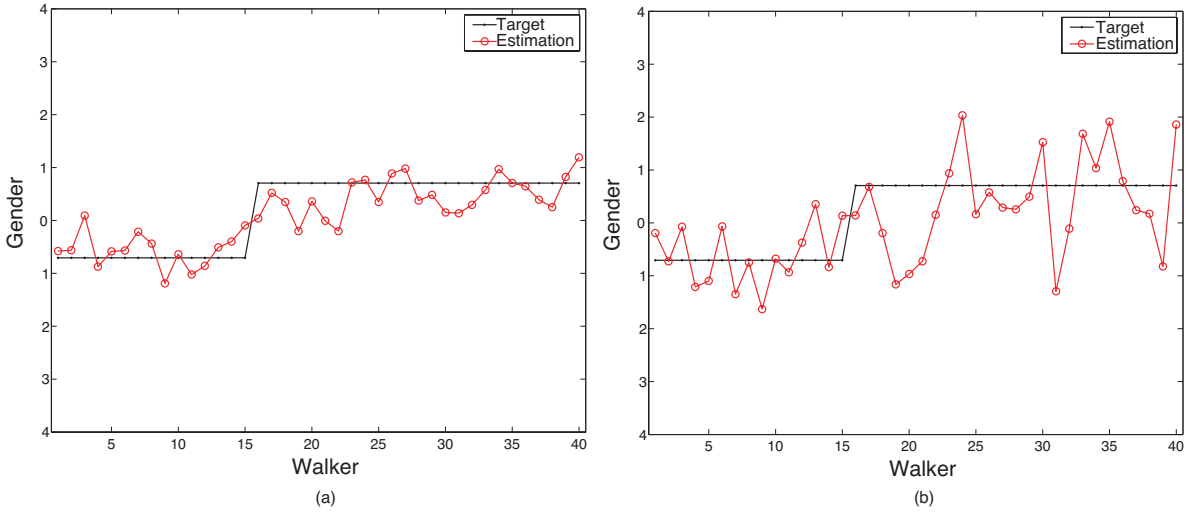

Figure 12. Perceptual gender parameter estimation results. (a). Expressive estimation. (b). Non-expressive SSE estimation. (walkers #1-15: perceived-female; #16-40: perceived-male)

pressive 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

$$Error = \frac{\sum_{k=1}^{K} \omega_k \left[sign(\bar{g}_k) \stackrel{?}{=} sign(\hat{g}_k) \right]}{\sum_{k=1}^{K} \omega_k}. \qquad (22)$$

Using the new perceptual gradient function and the weighted error calculation in the cross-validation technique, the single optimal expressive model was computed (at a 60% modal fit) and tested on the perceptually-labeled walking data. The resulting expressive model produced a weighted classification rate of 95.5% for all 40 walkers (the corresponding optimal SSE model produced a weighted classification rate of 88%).

Summary and conclusion

We presented an approach for gender recognition of point-light walkers using an expressive three-mode principal components framework. The approach initially factors prototype female and male walkers into their three-mode principal components to provide individual basis sets for the body posture, temporal trajectories, and gender changes. The main advantage of this multi-modal basis set is that it offers a low-dimensional decomposition of the data suitable for incorporating expressive weights on trajectories to bias the model estimation of gender. The method automatically learns the expressive feature weights from labeled training data, and therefore can adapt to different recognition contexts for the same underlying data.

We presented two types of gender labeling of the training data to learn the values of the expressive weights. Physical labeling assigns the true physical gender to each walker (*is* female/male). Perceptual labeling assigns genders resulting from a perceptual classification task to attain the observed gender (*appears* female/male). The labeled training data are used in a gradient descent-learning algorithm to solve for the expressive weight values needed to bias the model estimation of gender to the desired training values. Instead of matching a new walker to several examples for recognition, our expressive model is used to directly compute a gender value/label for the walker.

Results using 40 walkers (20 female, 20 male) labeled with their physical gender and our expressive model showed a recognition rate of 92.5%. These results are consistent with the recognition rate reported in Troje (2002) using a two-stage PCA and a linear classifier on a similar dataset of 40 walkers. We also trained the model using perceptually-based labels for the walkers. We first conducted a gender classification experiment using 15 observers of the 40 walkers. Results showed a 69% classification rate by human observers. We averaged the 15 gender selections for each walker to form a consistency value, where the thresholded consistency was used to assign the dominant perceived gender label to each walker for training the expressive model. Using the trained expressive model (on the perceptually-labeled data), a recognition rate of 90% was achieved. A model trained with consistency-weighted examples produced a higher 95.5% weighted classification rate. The results demonstrate that our approach can successfully and automatically adapt to different gender contexts (physical and perceptual), and that it can outperform a standard non-expressive SSE model (with no expressive weights).

Because gender recognition has been an active research domain for several years, we feel our model has merit for further analysis in this area. We plan to examine other representations (e.g., joint-angles), which may result in learned weight values that are highly correlated with higher-level movement interpretations. Then our model may help to provide insight to which features may be used by human observers during the gender classification task.

Appendix A: Singular value decomposition

Singular value decomposition (SVD) (Strang, 1993) can be used to factorize any $m \times n$ matrix $\mathbf{A}$ into

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}} = \underbrace{\begin{bmatrix} | & & | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_r & \cdots & \mathbf{u}_m \\ | & & | & & | \end{bmatrix}}_{m \times m \text{ (colomn space)}} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}}_{m \times n \text{ (singular values)}} \underbrace{\begin{bmatrix} - & \mathbf{v}_1 & - \\ & \vdots & \\ - & \mathbf{v}_r & - \end{bmatrix}}_{n \times n \text{ (row space)}},$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices and $\mathbf{\Sigma}$ is diagonal with r singular values $\sigma_1,\ldots,\sigma_r$. The columns of $\mathbf{U}$ correspond to a column space of $\mathbf{A}$, where any column of $\mathbf{A}$ can be formed by a linear combination of the columns in $\mathbf{U}$. Similarly, the rows of $\mathbf{V}^{\mathrm{T}}$ correspond to a row space of $\mathbf{A}$, where any row in $\mathbf{A}$ can be constructed by a linear combination of the rows in $\mathbf{V}^{\mathrm{T}}$.

The matrices $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{\Sigma}$ can be computed from the eigenvectors and eigenvalues of $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$, as shown by

$$\mathbf{A}\mathbf{A}^{\mathrm{T}} = (\mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}})(\mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}})^{\mathrm{T}} = \mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}}\mathbf{V}\Sigma^{\mathrm{T}}\mathbf{U}^{\mathrm{T}} = \mathbf{U}\Sigma^2\mathbf{U}^{\mathrm{T}}$$

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = (\mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}})^{\mathrm{T}}(\mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}}) = \mathbf{V}\Sigma^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}} = \mathbf{V}\Sigma^2\mathbf{V}^{\mathrm{T}},$$

where $\mathbf{U}$ is the orthonormal eigenvector basis of $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ and $\mathbf{V}$ is the orthonormal eigenvector basis of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$. The square of the singular values σ_i correspond to the eigenvalues.

The three-mode factorization of the prototype gender data $\bar{\mathbf{Z}}$ decomposes it into three orthonormal matrices $\mathbf{P}$, $\mathbf{T}$, and $\mathbf{S}$ that span the column (posture), row (time), and slice (gender) dimensions of the cube (see Figure 1c). The

desired basis sets can be found with SVD using three different 2D matrix-flattening arrangements of $\bar{\mathbf{Z}}$

$$\text{Posture: } \mathbf{P} = \text{columnSpace}([\, \bar{\mathbf{Z}}_f \mid \bar{\mathbf{Z}}_m \,])$$

$$\text{Time: } \mathbf{T} = \text{columnSpace}([\, \bar{\mathbf{Z}}_f^{\mathrm{T}} \mid \bar{\mathbf{Z}}_m^{\mathrm{T}} \,])$$

$$\text{Gender: } \mathbf{G} = \text{rowSpace}([\, \vec{\mathbf{Z}}_f \mid \vec{\mathbf{Z}}_m \,]),$$

where $\bar{\mathbf{Z}}_{\{f,m\}}^{\mathrm{T}}$ is the transpose of $\bar{\mathbf{Z}}_{\{f,m\}}$, and $\vec{\mathbf{Z}}_{\{f,m\}}$ is the rasterized column vector of matrix $\bar{\mathbf{Z}}_{\{f,m\}}$ (concatenation of point-light trajectories for each gender into a single column vector), and $[\mathbf{X} \mid \mathbf{Y}]$ is a matrix with the columns of $\mathbf{X}$ followed by the columns of $\mathbf{Y}$. Note that no two of the three basis sets can be produced within a single two-mode (matrix) factorization of $\bar{\mathbf{Z}}$.

Acknowledgments

This research was supported by National Science Foundation Grant No. 0236653 and the 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 data used in these experiments.

Commercial relationships: none.
Corresponding author: James W. Davis.
Email: jwdavis@cis.ohio-state.edu.
Address: Dept. of Computer and Information Science,
491 Dreese Laboratories, 2015 Neil Avenue,
Ohio State University, Columbus, OH 43210.

References

Alexa, M., & Muller, W. (2000). Representing animations by principal components. *Computer Graphics Forum*, *19*(3), 411-418.

Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and spatial actors in gait perception that influence gender recognition. *Perception & Psychophysics*, *23*(2), 145–152. [PubMed]

Beardsworth, T., & Buckner, T. (1981). The ability to recognize oneself from a video recording of one's movements without seeing one's body. *Bulletin of the Psychonomic Society*, *18*, 19–22.

Bingham, G. P. (1987). Kinematic form and scaling: Further investigations on the visual perception of lifted weight. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(2), 155–177. [PubMed]

Bingham, G. P. (1993). Scaling judgments of lifted weight: Lifter size and the role of the standard. *Ecological Psychology*, *5*(1), 31–64.

Black, M., Yacoob, Y., Jepson, A., & Fleet, D. (1997). Learning parameterized models of image motion. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 561–567.

Bobick, A., & Davis, J. W. (1996). An appearance-based representation of action. In *Proceedings of International Conference on Pattern Recognition*, 307–312.

Brand, M., & Hertzmann, A. (2000). Style machines. In *Proceedings of ACM SIGGRAPH*, 183–192.

Brand, M. (2001). Morphable 3D models from video. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 456–463.

Brownlow, S., Dixon, A., Egbert, C., & Radcliffe, R. (1997). Perception of movement and dancer characteristics from point-light displays of dance. *Psychological Record*, *47*, 411–421.

Burden, R., & Faires, J. (1993). *Numerical analysis*. Boston: PWS.

Chi, D., Costa, M., Zhao, L., & Badler, N. (2000). The EMOTE model for effort and shape. In *Proceedings of ACM SIGGRAPH*, 173–182.

Crawley, R., Good, J., Still, A., & Valenti, S. (2000). Perception of sex from complex body movement in young children. *Ecological Psychology*, *12*, 231–240.

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familarity cues. *Bulletin of the Psychonomic Society*, *9*, 353–356.

Cutting, J. E., Proffitt, D. R., & Kozlowski, L. T. (1978). A biomechanical invariant for gait perception. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(3), 357–372. [PubMed]

Davis, J. W. (2001). Visual categorization of children and adult walking styles. In *Proceedings of International Conference on Audio- and Video-based Biometric Person Authentication*, 295–300.

Davis, J. W., & Gao, H. (2003a). An expressive three-mode principal components model of human action style. *Image and Vision Computing*, *21*(11), 1001–1016.

Davis, J. W., & Gao, H. (2003b). Recognizing human action efforts: An adaptive three-mode PCA framework. In *Proceedings of IEEE International Conference on Computer Vision*, 1463–1469.

Davis, J. W., Gao, H., & Kannappan, V. (2002). A three-mode expressive feature model of action effort. In *Proceedings of IEEE Workshop on Motion and Video Computing*, 139–144.

Davis, J. W., & Kannappan, V. (2002). Expressive features for movement exaggeration. In *ACM SIGGRAPH Conference Abstracts and Applications*, 182.

Davis, J. W., & Taylor, S. (2002). Analysis and recognition of walking movements. In *Proceedings of International Conference on Pattern Recognition*, 315–318.

Dittrich, W. H., Troscianko, T., Lea, S. E., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, *25*(6), 727–738. [PubMed]

Giese, M., & Poggio, T. (2000). Morphable models for the analysis and synthesis of complex motion patterns. *International Journal of Computer Vision*, *38*(1), 59–73.

Hill, H., & Johnston, A. (2001) Categorizing sex and identity from the biological motion of faces. *Current Biology*, *11*(11), 880–885. [PubMed]

Hill, H., & Pollick, F. E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light displays. *Psychological Science*, *11*(3), 223–228. [PubMed]

Hirashima, S. (1999). Recognition on the gender of point-light walkers moving in different directions. *Japanese Journal of Psychology*, *70*(2), 149–153.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*(2), 201–211.

Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from dynamic point-light display. *Perception & Psychophysics*, *21*(6), 575–580.

Kozlowski, L. T., & Cutting, J. E. (1978). Recognizing the gender of walkers from point-lights mounted on ankles: Some second thoughts. *Perception & Psychophysics*, *23*(5), 459.

Kroonenberg, P. (1983). *Three-mode principal component analysis theory and applications*. Leiden: DSWO Press.

Kroonenberg, P., & Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, *45*(1), 69–97.

Li, N., Dettmer, S., & Shah, M. (1997). Visually recognizing speech using eigensequences. In M. Shah & R. Jain (Eds.), *Motion-based recognition* (pp. 345–371). Amsterdam: Kluwer Academic Publishing.

Mason, C. R., Gomez, J. E., & Ebner, T. J. (2001). Hand synergies during reach-to-grasp. *Journal of Neurophysiology*, *86*(6), 2896–2910. [PubMed]

Mather, G., & Murdoch, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *258*, 273–279.

Montepare, J. M., & Zebrowitz-McArthur, L. (1988). Impressions of people created by age-related qualities of their gaits. *Journal of Personality and Social Psychology*, *55*(4), 547–556. [PubMed]

Murase, H., & Nayar, S. (1995). Visual learning and recognition of 3D objects from appearance. *International Journal of Computer Vision*, *14*(1), 5–24.

Murray, M. P., Kory, R. C., & Sepic, S. B. (1970). Walking patterns of normal women. *Archives of Physical Medicine and Rehabilitation*, *51*, 637–650. [PubMed]

Pollick, F. E., Lestou, V., Ryu, J., & Cho, S. B. (2002). Estimating the efficiency of recognizing gender and affect from biological motion. *Vision Research*, *42*(20), 2345–2355. [PubMed]

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Stanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, *82*(2), B51–B61. [PubMed]

Runeson, S., & Frykholm, G. (1981). Visual perception of lifted weight. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(4), 733–740. [PubMed]

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition and deceptive intention. *Journal of Experimental Psychology: General*, *112*(4), 585–615.

Strang, G. (1993). *Introduction to linear algebra*. Wellesley: Wellesley-Cambridge Press.

Tenenbaum, J., & Freeman, W. (1997). Separating style and content. In M. Mozer, M. Jordan, & T. Petsche (Eds.), *Advances in neural information processing systems* (Vol. 9, p. 662). Boston: MIT Press.

Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, *2*(5), 371–387, http://journalofvision.org/2/5/2/, doi:10.1167/2.5.2. [PubMed] [Article]

Tucker, L. (1966). Some mathematical notes on three-mode factor analysis. *Psychometricka*, *31*(3), 279–311.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, *3*(1), 71–86.

Unuma, M., Anjyo, K., & Takeuchi, R. (1995). Fourier principles for emotion-based human figure animation. In *Proceedings of ACM SIGGRAPH*, 91–96.

Vasilescu, M. (2001). Human motion signatures for character animation. In *ACM SIGGRAPH Conference Abstracts and Applications*, 200.

Vasilescu, M., & Terzopoulos, D. (2002). Multilinear analysis of image ensembles: TensorFaces. In *Proceedings of European Conference on Computer Vision*, 447–460.

Walk, R., & Homan, C. (1984). Emotion and dance in dynamic light displays. *Bulletin of the Psychonomic Society*, *22*, 437–440.

Wilson, A., & Bobick, A. (1999). Parametric Hidden Markov Models for gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*(9), 884–900.

Yacoob, Y., & Black, M. (1999). Parameterized modeling and recognition of activities. *Computer Vision and Image Understanding*, *73*(2), 232–247.

Yamamoto, M., Kondo, T., Yamagiwa, T., & Yamanaka, K. (1998). Skill recognition. In *Proceedings of International Conference on Automatic Face and Gesture Recognition*, 604–609.