Multilinear modeling for robust identity recognition from gait

Fabio Cuzzolin
Oxford Brookes University
Oxford, UK

Abstract

Human identification from gait is a challenging task in realistic surveillance scenarios in which people walking along arbitrary directions are viewed by a single camera. However, viewpoint is only one of the many covariate factors limiting the efficacy of gait recognition as a reliable biometric. In this chapter we address the problem of robust identity recognition in the framework of multilinear models. Bilinear models, in particular, allow us to classify the "content" of human motions of unknown "style" (covariate factor). We illustrate a three-layer scheme in which image sequences are first mapped to observation vectors of fixed dimension using Markov modeling, to be later classified by an asymmetric bilinear model. We show tests on the CMU Mobo database that prove that bilinear separation outperforms other common approaches, allowing robust view- and action-invariant identity recognition. Finally, we give an overview of the available tensor factorization techniques, and outline their potential applications to gait recognition. The design of algorithms insensitive to multiple covariate factors is in sight.

Introduction

Biometrics has received a growing attention in the last decade, as automatic identification systems for surveillance and security started to enjoy widespread diffusion. Biometrics like face, iris, or fingerprint recognition, in particular, have been thoroughly employed. They suffer, however, from two major limitations: they cannot be used at a distance, and require user cooperation. Such assumptions are simply not sensible in real-world scenarios, e.g. surveillance of public areas.

Interestingly, psychological studies show that people are capable of recognizing their friends just from the way they walk, even when their "gait" is poorly represented by point light display (Cutting & Kozlowski, 1977). Gait has several advantages over other biometrics, as it can be measured at a distance, is difficult to disguise or occlude, and can be identified even in low-resolution images. Most importantly gait recognition is *non-cooperative* in nature. The person to identify can move freely in the surveyed environment, and is possibly unaware of his/her identity being checked.

The problem of recognizing people from natural gait has been studied by several researchers (Gafurov, 2007; Nixon & Carter, 2006), starting from a seminal work of Niyogi and Adelson (1994). Gait analysis can also be applied to gender recognition (Li et al., 2008), as different pieces of information like gender or emotion are contained in a walking gait and can be recognized. Abnormalities of gait patterns for the diagnosis of certain diseases can also be automatically detected (Wang, 2006). Furthermore, gait and face biometrics can be easily integrated for human identity recognition (Zhou & Bhanu, 2007; Jafri & Arabnia, 2008).

Influence of covariates

Despite of its attractive features, though, gait identification is still far from being ready to be deployed in practice.

What limits the adoption of gait recognition systems in real-world scenarios is the influence of a large number of so-called “covariate” factors which affect appearance and dynamics of the gait. These include walking surface, lightning, camera setup (viewpoint), but also footwear and clothing, carrying conditions, time of execution, walking speed.
The correlation between those factors can be indeed very significant as pointed out in (Li et al., 2008), making gait difficult to measure and classify.
In the last few years a number of public databases have been made available and can be used as a common ground to validate the variety of algorithms that have been proposed. The USF database (Sarkar et al., 2005), for instance, was specifically designed to study the effect of covariate factors on identity classification in a realistic, outdoor context with cameras located at a distance.

View-invariance
The most important of those covariate factors is probably viewpoint variation. In the USF database, however, experiments contemplate only two cameras at fairly close viewpoints (with a separation of some 30 degrees). Also people are shot while walking along the opposite side of an ellipse: the resulting views are almost fronto-parallel. As a result appearance-based algorithms work well in the reported experiments concerning viewpoint variability, while one would expect them to perform poorly for widely separated views.
In a realistic setup, the person to identify steps into the surveyed area from an arbitrary direction. View-invariance (Urtasun & Fua, 2004; Yam et al., 2004; Bhanu & Han, 2002; Kale et al., 2003; Shakhnarovich et al., 2001; Johnson & Bobick, 2001) is then a crucial issue to make identification from gait suitable for real-world applications.
This problem has actually been studied in the gait ID context by many groups (Han et al., 2005). If a 3D articulated model of the moving person is available, tracking can be used as a pre-processing stage to drive recognition. Cunado et al. (1999), for instance, used their evidence gathering technique to analyze the leg motion in both walking and running gait. Yam et al. (2004) also worked on a similar model-based approach. Urtasun and Fua (2004) proposed an approach to gait analysis that relies on fitting 3D temporal motion models to synchronized video sequences. Bhanu and Han (2002) matched a 3D kinematic model to 2D silhouettes. Viewpoint invariance is achieved in (Spencer & Carter, 2002) by means of a hip/leg model, including camera elevation angle as an additional parameter.
Model-based 3D tracking, however, is a difficult task. Manual initialization of the model is often required, while optimization in a higher-dimensional parameter space suffers from convergence issues. Kale et al. (2003) proposed as an alternative a method for generating a synthetic side-view of the moving person using a single camera, if the person is far enough. Shakhnarovich et al. (2001) suggested a view-normalization technique in a multiple camera framework, using the volumetric intersection of the visual hulls of all camera silhouettes. A 3D model is also set up in (Zhao et al., 2006) using sequences acquired by multiple cameras, so that the length of key limbs and their motion trajectories can be extracted and recognized. Johnson and Bobick (2001) presented a multi-view gait recognition method using static body parameters recovered during the walking motion across multiple views.
More recently, Rogez et al. (2006) used the structure of man-made environments to transform the available image(s) to frontal views, while Makihara et al. (2006) proposed a view transformation model in the frequency domain, acting on features obtained by Fourier analysis of a spatiotemporal volume.
An approach to multiple view fusion based on the “product of sum” rule was proposed in (Lu and Zhang, 2007). Different features and classification methods were there compared. The discriminating power of different views was analyzed in (Huang & Boulgouris, 2008). Several evidence combination methods were tested on the CMU Mobo database (Gross &

Shi, 2001).
More in general, the effects of all the different covariates have not been yet thoroughly investigated, even though some effort has been recently done is this direction. Bouchrika and Nixon (2008) conducted a comparative study of their influence in gait analysis. Veres et al. (2005) proposed a remarkable predictive model of the "time of execution" covariate to improve recognition performance. The issue has however been approached so far on an empirical basis, i.e. by trying to measure the influence of individual covariate factors. A principled strategy for their treatment has not yet been brought forward.

Chapter's objectives
A general framework for addressing the issue of covariate factors in gait recognition is provided by *multilinear* or *tensorial models*. These are mathematical descriptions of the way different factors *linearly* interacts in a mixed training set, yielding the walking gaits we actually observe.
The problem of recovering those factors is often referred to in the literature as *nonnegative tensor factorization* or *NTF* (Tao, 2006). The PARAFAC model for multi-way analysis (Kiers, 2000) was first introduced for continuous electroencephalogram (EEG) classification in the context of brain-computer interfaces (Morup et al., 2006). A different multi-layer method for 3D NTF was proposed by Cichocki et al. (2007). Porteus et al. (2008) introduced a generative Bayesian probabilistic model for unsupervised tensor factorization. It consists of several interacting LDA models, one for each modality (factor), corredated with a Gibbs sampler for inference. Other approaches to NTF can be found in recent papers like
(Lee et al., 2007; Shashua & Hazan, 2005; Boutsidis et al., 2006).
Bilinear models in particular (Tenenbaum & Freeman, 2000) are the best studied and probably the most popular among multilinear models. They can be seen as a tool for separating two factors, usually called "style" and "content" of the objects to classify. This way they allow (for instance) to build a classifier which, given a new sequence in which a *known* person is seen from a view *not* in the training set, can iteratively estimate both identity and view parameters, significantly improving recognition performances.
In this chapter we propose a *three-layer model* in which each motion sequence is considered as an observation depending on three factors (*identity*, *action* type, and *view*). A bilinear model can be trained from those observations by considering two such factors at a time. While in the first layer features are extracted from single images, in the second stage each feature sequence is given as input to a hidden Markov model (HMM). Assuming fixed dynamics, the HMM clusters the sequence into a fixed number of poses. The stacked vector of these poses eventually forms a vector which represents the input motion. After learning a bilinear model for such set of observation vectors we can then classify (determine the content of) new sequences characterized by a different style label.
We illustrate experiments on the CMU Mobo database on view-invariant and action invariant identity recognition. They clearly show how this approach performs significantly better than other standard gait recognition algorithms.
To conclude we outline several possible natural extensions of this methodology to multilinear modeling, in the perspective of providing a comprehensive framework for dealing in a consistent way with an arbitrary number of covariates.

Bilinear models

Bilinear models were introduced by Tenenbaum & Freeman (2000) as a tool for separating what they called "style" and "content" of a set of objects to classify, i.e., two distinct class labels $s \in [1,...,S]$ and $c \in [1,...,C]$ attributed to such objects. Common but useful examples are

font and alphabet letter in writing, or word and accent in speaking.
Consider a training set of K-dimensional observations $\boldsymbol{y}_k^{sc}$, $k = 1,...,K$ characterized by a style s and a content c, both represented as parameter vectors $\boldsymbol{a}^s$ and $\boldsymbol{b}^c$ of dimension I and J respectively. In the *symmetric* model we assume that these observations can be written as

$$\boldsymbol{y}_k^{sc} = \sum_{i=1}^{I} \sum_{j=1}^{J} w_{ijk} a_i^s b_j^c \tag{1}$$

where a_i^s and b_j^c are the scalar components of the vectors $\boldsymbol{a}^s$ and $\boldsymbol{b}^c$ respectively.
Let $\boldsymbol{W}_k$ denote the k-th matrix of dimension $I \times J$ with entries w_{ijk}. The symmetric model (1) can then be rewritten as

$$\boldsymbol{y}^{sc}{}_k = (\boldsymbol{a}^s)^T \boldsymbol{W}_k \boldsymbol{b}^c \tag{2}$$

where ^{T} denotes the transpose of a matrix or vector. The K matrices $\boldsymbol{W}_k$, $k = 1,...,K$ define a *bilinear map* from the style and content spaces to the K-dimensional observation space.
When the interaction factors can vary with style (i.e. w^s_{ijk} depends on s) we get an *asymmetric* model

$$\boldsymbol{y}^{sc} = \boldsymbol{A}^s \boldsymbol{b}^c. \tag{3}$$

Here $\boldsymbol{A}^s$ denotes the $K \times J$ matrix with entries $\{a^s_{jk} = \sum_i w^s_{ijk} a^s_i\}$, a *style-specific linear map* from the content space to the observation space (see Figure 1-right).

Training an asymmetric model
A bilinear model can be fit to a training set of observations endowed with two labels by means of simple linear algebraic techniques. If the training set has (roughly) the same number of measurements $\boldsymbol{y}^{sc}$ for each style and each content class we can use classical singular value decomposition (SVD). If we stack the training data into the $(SK) \times C$ matrix

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}^{11} & \dots & \boldsymbol{y}^{1C} \\ \dots & \dots & \dots \\ \boldsymbol{y}^{S1} & \dots & \boldsymbol{y}^{SC} \end{bmatrix} \tag{4}$$

the asymmetric model can be written as $\boldsymbol{Y} = \boldsymbol{AB}$ where $\boldsymbol{A}$ and $\boldsymbol{B}$ are the stacked style and content parameter matrices, $\boldsymbol{A} = [\boldsymbol{A}^1 \dots \boldsymbol{A}^S]^T$, $\boldsymbol{B} = [\boldsymbol{b}^1 \dots \boldsymbol{b}^C]$.
The least-square optimal style and content parameters are then easily found by computing the SVD of (4), $\boldsymbol{Y} = \boldsymbol{USV}^T$, and assigning

$$\boldsymbol{A} = [\boldsymbol{US}]_{col=1..J}, \quad \boldsymbol{B} = [\boldsymbol{V}^T]_{row=1..J}. \tag{5}$$

If the training data are not equally distributed among all the classes, a least-square optimum has to be found (Tenenbaum & Freeman, 2000).

Content classification of unknown style
Suppose that we have learnt a bilinear model from a training set of data. Suppose also that a new set of observations becomes available *in a new style*, different from all those already known in the training set, but with *content* labels *among those learned in advance*. In this case an iterative procedure can be set up to factor out the effect of style and classify the content

labels of the new observations.
Notice that if we know the content class assignments of the new data we can find the parameters for the new style s' by solving for $\boldsymbol{A}^{s'}$ in the asymmetric model (3). Analogously, having a map $\boldsymbol{A}^{s'}$ for the new style we can easily classify the new "test" vectors $\boldsymbol{y}$ by measuring their distance $\|\boldsymbol{y} - \boldsymbol{A}^{s'}\boldsymbol{b}^c\|$ from $\boldsymbol{A}^{s'}\boldsymbol{b}^c$ for each (known) content vector $\boldsymbol{b}^c$.
The issue can be solved by fitting a mixture model to the learnt bilinear model by means of the EM algorithm (Dempster et al., 1977). The EM algorithm alternates between computing the probabilities $p(c|s')$ of the current content label given an estimate s' of the style (E step), and estimating a linear map $\boldsymbol{A}^{s'}$ for the unknown style s' given the current content class probabilities $p(c|s')$ (M step).
We assume that the probability of observing a measurement $\mathbf{y}$ given the new style s' and a content label c is given by a Gaussian distribution

$$p(\boldsymbol{y}|s',c) = \exp\left(\frac{-\|\boldsymbol{y} - \boldsymbol{A}^{s'}\boldsymbol{b}^c\|^2}{2\sigma^2}\right). \tag{6}$$

The total probability of such an observation $\boldsymbol{y}$ (notice that the general formulation allows for the presence of more than one unknown style, (Tenenbaum & Freeman, 2000)) is then

$$p(\boldsymbol{y}) = \sum_c p(\boldsymbol{y}|s',c)\, p(s',c) \tag{7}$$

where in absence of prior information $p(s',c)$ is supposed to be equally distributed.
In the E step the algorithm computes the joint probability of the labels given the data

$$p(s',c|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|s',c)\, p(s',c)}{p(\boldsymbol{y})} \tag{8}$$

(using Bayes' rule) and classifies the test data by finding the content class c which maximizes $p(c|\boldsymbol{y}) = p(s',c|\boldsymbol{y})$.
In the M step the style matrix $\boldsymbol{A}^{s'}$ which maximizes the log-likelihood of the test data is estimated. This yields

$$\boldsymbol{A}^{s'} = \frac{\sum_c \boldsymbol{m}^{s'c}(\boldsymbol{b}^c)^T}{\sum_c n^{s'c}\boldsymbol{b}^c(\boldsymbol{b}^c)^T}, \tag{9}$$

where $\boldsymbol{m}^{s'c} = \sum_y p(s',c|\boldsymbol{y})\boldsymbol{y}$ is the mean observation weighted by the probability of having style s' and content c, and $n^{s'c} = \sum_y p(s',c|\boldsymbol{y})$ is a normalization factor.
The effectiveness of the method critically depends on whether the observation vectors actually meet the assumption of bilinearity. However, it was originally presented as a way of finding *approximate* solutions to problems in which two factors are involved, without precise context-based knowledge, and that is the way it is used here.

A three-layer model

In human motion analysis movements (and walking gaits in particular) can be characterized by a number of different labels. They can be classified according to the identity of the moving person, their emotional state, the category of action performed (i.e. walking, reaching out, pointing, etc.), or (if the number of cameras is finite) the viewpoint from which the sequence

is shot.
As a matter of fact each covariate factor can be seen as an additional label assigned to each walking gait sequence. Covariate-free gait recognition can then be naturally formulated in terms of multilinear modeling (Elgammal and Lee, 2004).
In this chapter we illustrate the use of bilinear models to represent and classify gaits regardless the "style" with which they are executed, i.e., the value of the (in this case single) covariate factor. In practice this allows us to address problems like *view-invariant identity recognition*, identity recognition from *unknown* gaits, and to ensure robustness with respect to emotional state, clothing, elapsed time, etcetera.
We propose a *three-layer model* in which each motion sequence is considered as an observation which depends on all covariate factors. A bilinear model can be trained by considering two of those factors at a time. We can then apply bilinear classification to recognize gaits regardless their style.

First layer: feature representation
In gait ID images are usually preprocessed in order to extract the silhouettes of the walking person. We chose a simple but effective way of computing feature measurements from each such silhouette. More precisely, we detect its center of mass, rescale it to the corresponding bounding box, and project its contours on to one or more lines passing through its barycenter (see Figure 1-right). We favored this approach after testing a number of other different representations: the principal axes of the body-parts as they appear in the image (Lee & Grimson, 2002), size functions (Frosini, 1991), and a PCA-based representation of the contours. All turned out to be rather unstable.

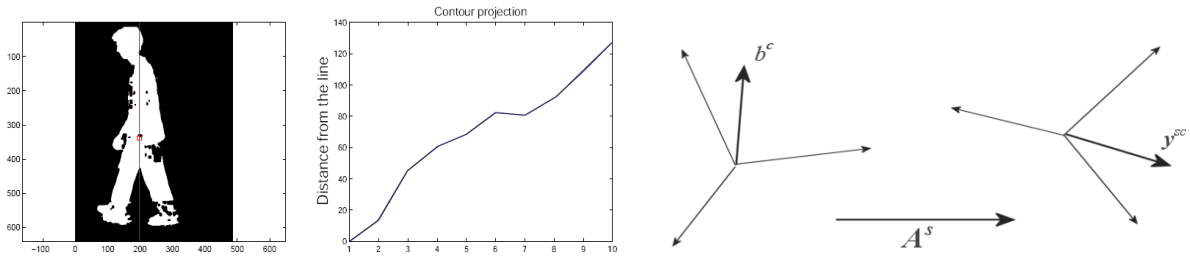

Figure 1: Left*: Feature extraction. First a number of lines passing through the center of mass of the silhouette are selected. Then for each such line the distance of the points on the contour of the silhouette from it is computed (here the segment is sub-divided into 10 intervals). The collection of all such distance values for all lines eventually forms the feature vector representing the image. Right: bilinear modeling. Each observation* $\boldsymbol{y}^{sc}$ *is the result of applying a style-specific linear map* $\boldsymbol{A}^s$ *to a vector* $\boldsymbol{b}^c$ *of some abstract "content space".*

Second layer: HMMs as sequence descriptors
If the contour of the silhouette is projected onto 2 orthogonal lines passing through its barycenter, and we divide each line segment into 10 equally spaced intervals, each image ends up being represented by a 40-dimensional feature vector. Image sequences are then encoded as sequences of feature vectors, in general of different length (duration). To adapt them to their role of inputs for a bilinear model learning stage we need to transform those feature sequences into observation vectors *of the same size*.
Hidden Markov models (Elliot et al., 1995) provide us with such a tool.
Even though they have been widely applied to gesture or action recognition, HMMs have rarely been considered as a tool in a gait ID context (He & Debrunner, 2000; Sundaresan et al., 2003), mainly to describe (Kale et al., 2002, He & Debrunner, 2000) or normalize (Liu &

Sarkar, 2006) gait dynamics (Kale et al., 2004).

A *hidden Markov model* is a statistical model whose states $\{X_k\}$ form a *Markov chain.* The only observable quantity though is a corrupted version y_k of the state called *observation process*.
Using the notation in (Elliot et al., 1995) we can associate the elements of the finite state space $X = \{1,...,n\}$ with coordinate versors $e_i = (0,...,0,1,0,...,0) \in \mathrm{R}^n$ and write the model as

$$X_{k+1} = AX_k + V_{k+1}$$
$$y_{k+1} = CX_k + diag(W_{k+1})\Sigma X_k. \qquad (10)$$

Here $\{V_{k+1}\}$ is a sequence of martingale increments and $\{W_{k+1}\}$is a sequence of i.i.d. Gaussian noises $\mathcal{N}(0,1)$. Given a state $X_k = e_j$ the observations y_{k+1} are then assumed to have Gaussian distribution $p(y_{k+1}|X_k = e_j)$ centered around a vector c_j which corresponds to the j-th column of the matrix C.
The parameters of the hidden Markov model (10) are then the "transition matrix" $A = (a_{ij}) = \mathrm{P}(X_{k+1} = e_i|X_k = e_j)$, the matrix C collecting the means of the state-output distributions $p(y_{k+1}|X_k = e_j)$ and the matrix Σ of their variances. The matrices $A, C,$ and Σ can be estimated, given a sequence of observations $\{y_1,...,y_T\}$, using (again) the Expectation-Maximization (EM) algorithm (see (Elliot et al., 1995) for the details).

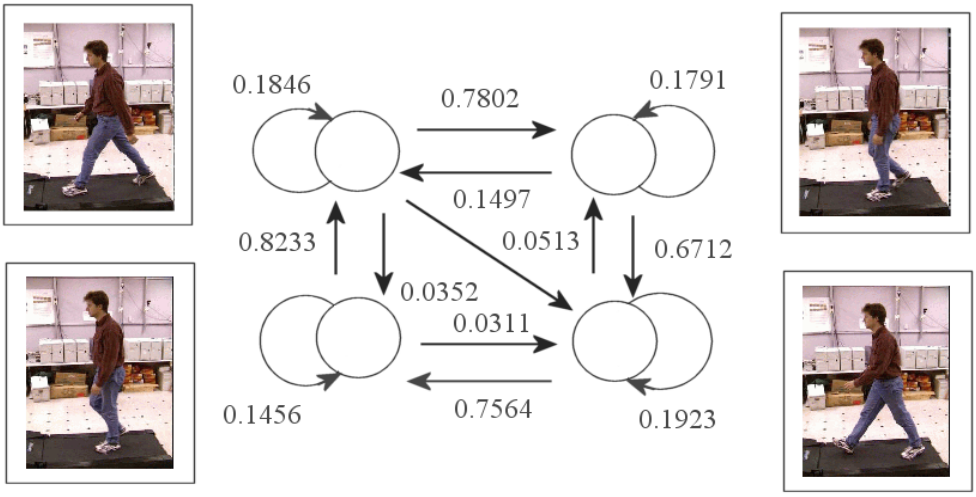

Figure 2. An example of hidden Markov model generated by a gait sequence. The HMM can be seen as a graph where each node represents a state (in this case N=4). Each state is associated with a key "pose" of the walking gait. Transitions between states are governed by the matrix A and are drawn as directed edges with attached a transition probability.

Let us now go back to the gait ID problem.
Given a sequence of feature vectors extracted from all the silhouettes of a sequence, EM yields as output a finite state representation (an HMM) of the motion. The latter is represented as a series of possible transitions (each associated with a certain probability) between key "poses" mathematically described by the states of the model (see Figure 2). The transition matrix A encodes the sequence's dynamics, while the columns of the C matrix represent the poses in the observation space.
In case of cyclic motions like the walking gait the dynamics is rather trivial, a circular series of transitions through the states of the HMM (see Figure 2 again). There is no need to

estimate the period of the cycle, as the poses are automatically associated with the states of the Markov model by the EM algorithm. For the same reason sequences with variable speed cause no trouble, in opposition to methods based on the estimation of the fundamental frequency of the motion (Little & Boyd, 1998).

Third layer: bilinear model of HMMs

Given the HMM which best fits the input feature sequence, its pose matrix *C* can be stacked into a single observation vector by simple concatenation of its columns.

If we select a fixed number *N* of poses for each sequence our training set of walking gaits can be encoded as a dataset of these observation vectors. They will have homogeneous size, even in the case in which the original sequences had different durations. Such vectors can later be used to build a bilinear model for the input training set of gait motions.

The procedure can then be summarized as follows:

- each training image sequence is mapped to a sequence of feature vectors;
- those feature sequences are fed to EM algorithm which delivers an *N*-state HMM for each training motion;
- the (pose) *C* matrix of each HMM is stacked to form a single observation vector;
- the algorithm recalled above is used to build an asymmetric bilinear model for the whole dataset.

The resulting three-layer model is depicted in Figure 3. Given a dataset of walking gaits, we can use this algorithm to built an asymmetric bilinear model from the sequences related to all style labels (covariate factors) but one. This will be our training set. We can then use the bilinear classifier to label the sequences associated with the remaining style (testing set).

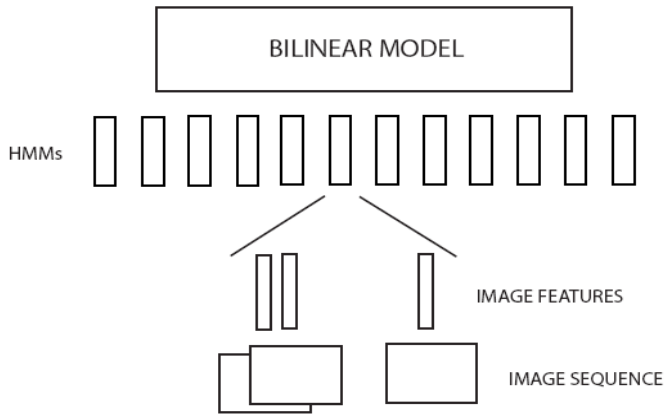

Figure 3: The proposed three-layer model. Features (bottom layer) are first extracted from each image of the sequence. The resulting feature sequences are fed to an HMM with a fixed number of states, yielding a dataset of Markov models, one for each sequence (second layer). The stacked versions of the (pose) C matrices of these models are then used as observation vectors to train an asymmetric bilinear model (top layer).

Experiments

We used the CMU Mobo database (Gross & Shi, 2001) to extensively test our bilinear approach to gait ID. As its six cameras are widely separated, Mobo gives us the chance of testing the algorithm in a rather realistic setup. In the database 25 different people perform four different walking-related actions: slow walk, fast walk, walking along an inclined slope, and walking while carrying a ball. The sequences were acquired indoor, with the subjects

walking on a treadmill at constant speed. The cameras are more or less equally spaced around the treadmill, roughly positioned around the origin of the world coordinate system (Gross & Shi, 2001). Each sequence is composed by some 340 frames, encompassing 9-10 full walking cycles. We renamed the six cameras originally called 3,5,7,13,16,17 as 1,2,3,4,5,6.

From view-invariant gait ID to ID-invariant action recognition

The video sequences of the Mobo database possess three different labels: identity, action, and viewpoint. Therefore we set up two series of tests in which asymmetric bilinear models were built by selecting identity as content label, and choosing a style label among the two remaining covariates. The two options were then: content=ID, style=view (*view-invariant gait ID*); content=ID, style=action (*action-invariant gait ID*).

The remaining factor was considered as a nuisance. Note that "action" here can be assimilated to classical covariates like walking surface (as the treadmill can be inclined or not) or carrying conditions (as the subject may or not carry a ball).

In each experiment we formed a different training set by considering the sequences related to all the style labels but one. We then built an asymmetric bilinear model as explained above. Eventually we used the sequences associated with the remaining style label as test data, and measured the performance of the bilinear classifier.

To get a fairly large dataset we adopted the period estimation technique of (Sarkar et al., 2005) to sub-divide the original long sequences into a larger number of subsequences, each spanning three walking cycles. We obtained a collection of 2080 sequences, almost equally distributed among the six views, the 25 IDs, and the four actions. We computed a feature matrix for each subsequence, and applied the HMM-EM algorithm with $N = 2$ states to generate a dataset of pose matrices C, each containing two pose vectors as columns. We finally stacked those columns into a single observation vector for each subsequence. These observation vectors would finally form our training set. We used the set of silhouettes provided with the database, after some preprocessing to remove small artifacts from the original images. In the following we report the performance of the algorithm using both the percentage of correct best matches and the percentage of test sequences for which the correct identity is one of the first three matches.

The bilinear classifier depends on a small number of parameters, in particular the variance σ of the mixture distribution (6) and the dimension J of the content space. They can be learnt in a preliminary stage by computing the score of the algorithm when applied to the training set for each value of the parameters. Basically the model needs a large enough content space to accommodate all the content labels. Most important is though the initial value of the probability $p(c|\boldsymbol{y})$ with which each test vector $\boldsymbol{y}$ belongs to a content class c. Again, this can be obtained from the training set by maximizing the classification performance, using some sort of *simulated annealing* technique to overcome local maxima.

View-invariant identity recognition

In the first series of tests we set "identity" as the content label and "viewpoint" as the style label (covariate). This way we could test the view-invariance of a gait ID bilinear classifier. We report here the results of different kinds of tests. In Figure 4 we selected the subset of the Mobo database associated with a single action (the nuisance, in this case). We then measured the performance of our bilinear classifier using view 1 as test view, for an increasing number of subjects (from 7 to 25). To get a flavor of the relative performance of our algorithm, we implemented a simple nearest neighbor classifier which assigns to each test sequence the identity of the closest Markov model. We measured distances between HMMs using the standard Kullback-Leibler divergence (Kullback & Leibler, 1951). Figure 4 clearly shows how the bilinear classifier greatly outperforms a naive NN classification of the Markov

models built from the gait sequences.

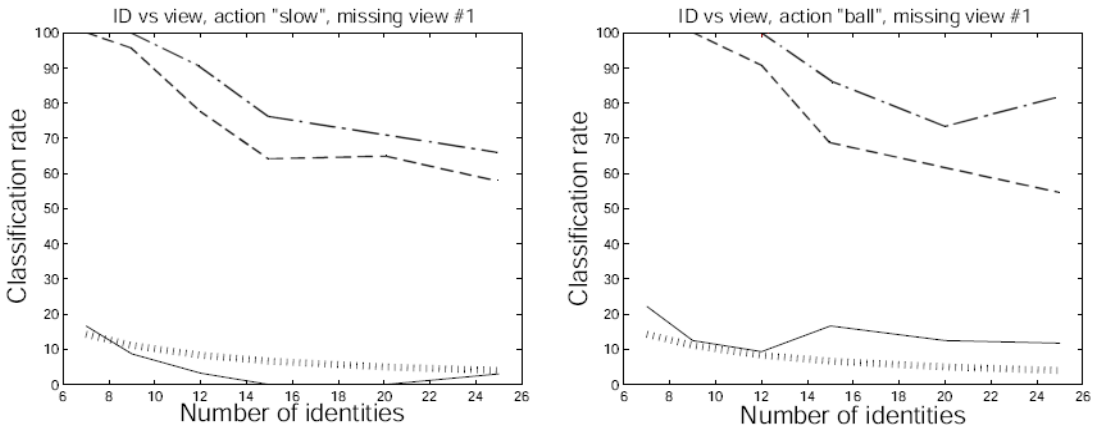

Figure 4: View-invariant gait ID for gait sequences related to the same action: "slow" (left) and "ball" (right). View 1 is used as the test view, while all the others are included in the training set. The classification rate is plotted versus an increasing number of subjects (from 7 to 25). The percentage of correct best matches is shown in dashed lines, while the rate of a correct match in the first 3 is plotted in dot-dashed lines. For comparison the performance of a KL-nearest neighbor classifier on the training set of HMMs is shown in solid black. As a reference pure chance is plotted using little vertical bars.

The depressing results of the KL-NN approach attest the difficulty of the task. You cannot just neglect the fact that image sequences come from widely separated viewpoints.

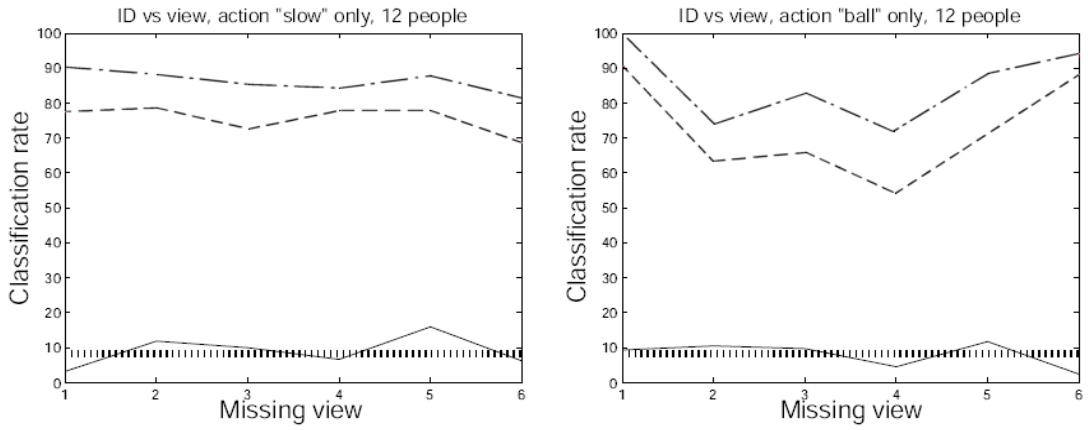

Figure 5: View-invariant gait ID for instances of the actions "slow" (left) and "ball" (right). The classification rate achieved for different test views (from 1 to 6) is plotted. Only the first 12 identities are here considered. Plot styles as above.

Figure 5 instead compares the two algorithms as the test viewpoint varies (from 1 to 6), for the two sub-datasets formed by instances of the actions "slow" and "ball", with 12 identities. Again the NN-KL classifier (which does *not* take into account the viewpoint from which is sequence is shot) performs around pure-chance levels. The bilinear classifier achieves instead excellent scores around 90% for some views. Relatively large variations in the second plot are due, in our opinion, to the parameter learning algorithm being stuck to a local maximum. Figure 6-left illustrates the performance of the algorithm as a function of the nuisance factor,

i.e. the performed action: ball=1, fast=2, incline=3, slow=4.

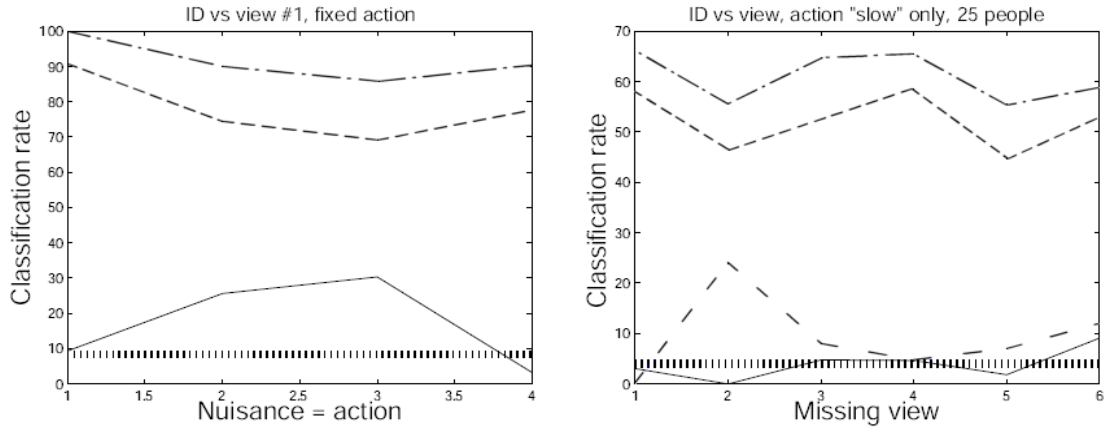

Figure 6: Performance of the bilinear classifier in view-invariant gait ID experiment. Left: Classification rate as a function of the nuisance (action), test view 1. Right: score for the dataset of sequences related to the action "slow", and different selection of the testview (from 1 to 6). This time all the 25 identities were considered. The classification rate of the baseline algorithm is the widely spaced dashed line in the right diagram: other line styles as above.

The classification rate of the bilinear classifier does not exhibit any particular dependence on the nuisance action. We also implemented for sake of comparison the *baseline algorithm* described in (Sarkar et al., 2005). The latter basically computes similarity scores between the test sequence *SP* and each training sequence *SG* by pairwise frame correlation. The baseline algorithm is used on the USF database to provide a performance reference.

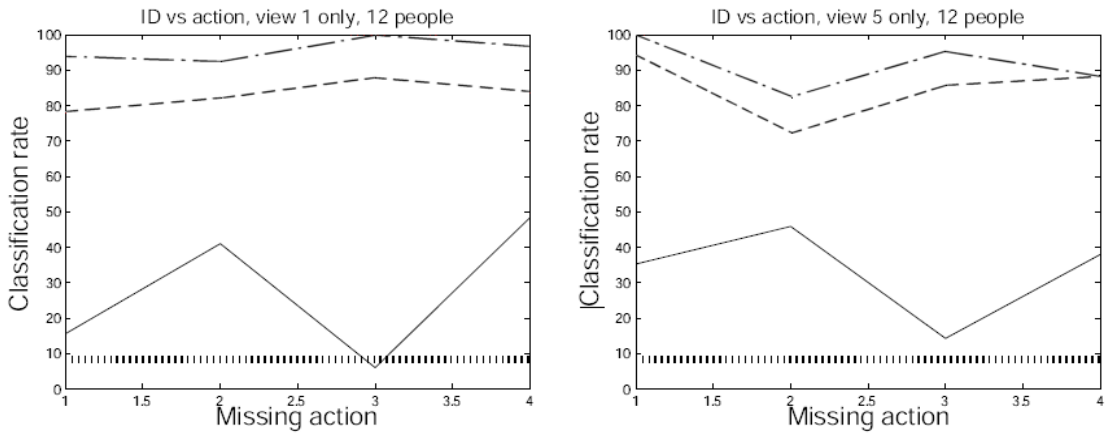

Figure 7: Action-invariant gait ID for sequences related to viewpoints 1 (left) and 5 (right). The classification rate is plotted versus different possible test actions (from 1 to 4). The first 12 identities were considered. Plot styles as in Figure 4.

Figure 6-right compares the results of bilinear classification with those of both the baseline algorithm and the KL-based approach for all the six possible test views, in the complete dataset comprising all 25 identities. The structure introduced by the bilinear model greatly improves the identification performance, rather homogeneously over all the views. The baseline algorithm instead seems to work better for sequences coming from cameras 2 and 3, which have rather close viewpoints, while it delivers the worst results for camera 1, the most

isolated from the others (Gross & Shi, 2001). The performance of the KL-based nearest neighbor approach is not distinguishable from pure chance.

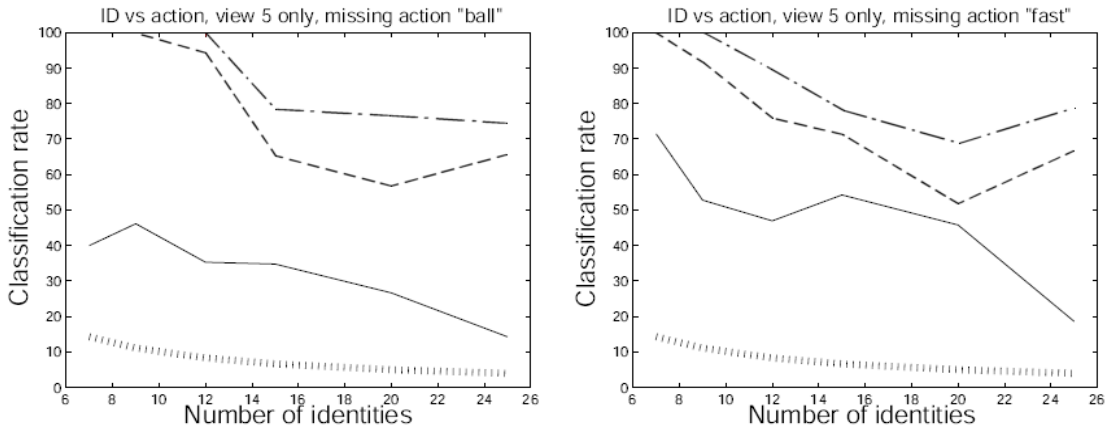

Figure 8: Action-invariant gait ID. In the left diagram sequences related to viewpoint (nuisance) #5 were considered, and "ball"was used as missing action (test style). In the right diagram sequences related to viewpoint #4 were considered, and "fast" used as test action. The classification rate is plotted versus an increasing number of subjects. Styles as above.

Action-invariant identity recognition

In a different experiment we validated the conjecture that a person can be recognized *even from an action he/she never performed,* provided that we have seen this action performed by other people in the past. In our case this assumption is quite reasonable, since all the actions in the database are nothing but variants of the gait gesture. Remember that some actions in the Mobo database correspond in fact to covariate factors like surface or carrying conditions. Here "1" denotes the action "slow", "2" denotes "fast", "3" stands for "walking on inclined slope", and "4" designates "walking while carrying a ball". We then built bilinear models for content=ID, style=action from a training set of sequences related to three actions, and classified the remaining sequences (instances of the fourth action) using the bilinear approach. Figures 7 and 8 support the ability of bilinear classification to allow identity recognition even from unknown gestures (or, equivalently, under different surface or carrying conditions, actions 3 and 4).

Figure 7 shows two diagrams in which identity recognition performances for sequences shot from viewpoints 1 (left) and 5 (right) only are selected, setting "action" as covariate factor (style). For all missing styles (actions) the three-stage bilinear classifier outperforms naive NN classification in the space of hidden Markov models. The performance of the latter is quite unstable, yielding different results for different unknown covariate values (actions), while bilinear classification appears to be quite consistent.

Figure 8 shows that the best-match ratio is around 90% for twelve persons, even though it slightly declines for a larger number of subjects (the parameter learning algorithm is stopped after a fixed period of time, yielding suboptimal models). The NN-KL classifier performs relatively better in this experiment, but well below an acceptable level.

Future developments: Extensions to multilinear modeling

The above experiments seem to prove that bilinear models are indeed capable of handling the influence of one covariate factor in gait recognition. In particular, we focused above on what is maybe the most important such factor, viewpoint. To provide a comprehensive framework

for covariate factor analysis, however, we need to extend our framework to *multilinear* models capable of handling many if not all the involved factors.
We can envisage two possible developments along this line. The first one concerns the compact representation of image sequences *as 3D tensors* instead of stacked column vectors.

Bilinear modeling of sequences as three-dimensional tensors
Reduction methods have been largely used to approach the gait recognition problem. Linear techniques in particular are very popular (Abdelkader et al., 2001; Murase & Sakai, 1996; Tolliver & Collins, 2003; Han & Bhanu, 2004). Ekinci et al., for instance (2007), applied to the problem Kernel PCA. An interesting biologically inspired work (Das et al., 2006) proposed a two-stage PCA to kinematic data to describe gait cycles.
Nonlinear dimensionality reduction has also been applied to the problem. Locally Linear Embedding was used in (Honggui & Xingguo, 2004) to detect gait cycles, with the shape of the embeddings providing the features. In (Kaziska & Srivastava, 2006) human gait was modeled and classified as a stochastic cyclostationary process on a nonlinear shape space.

Novel reduction methods which apply to *tensor* or *multilinear* data have also been recently investigated, yielding multilinear extensions of dimensionality reduction techniques like PCA. A *tensor* or n-mode matrix, is a higher order generalization of a vector (first order tensor) and a matrix (second order tensor). Formally, a tensor A of order N is a multilinear mapping over a set of vector spaces V_1, ..., V_N of dimensions I_1, ..., I_N. An element of A is denoted as $a_{i_1 \ldots i_n \ldots i_N}$ where $1 \leq i_n \leq I_n$.
In image analysis and computer vision inputs come naturally in the form of matrices (the images themselves) or third-order tensors (image sequences).
General tensor discriminant analysis was applied in (Tao et al., 2007) to three different image representations based on Gabor filters. Matrix-based dimensionality reduction was also applied in (Xu et al., 2006) to averaged silhouettes. A sophisticated application of marginal Fisher analysis on the result of tensor-based dimensionality reduction directly applied to grey-level images can instead be found in (Xu et al., 2007). Lu et al. (2006), on their side, proposed a multilinear PCA algorithm and applied it to gait analysis. They used a novel representation called EigenTensorGait in which each half cycle, seen as a third-order tensor is considered as one data sample. He et al. (2005) proposed a Tensor Subspace Analysis for second-order tensors (images) and compared their results with those produced by PCA and LDA.

A natural extension of the proposed three-layer framework will be the formulation of a model capable of handling observation sequences directly in the form of 3D tensors, instead of having to represent them as packed observation vectors. As learning and classification in bilinear models are implemented through SVD, this appears not to be an obstacle.

Multilinear covariate factor models
A real extension of the presented methodology to an arbitrary number of covariate factors, though, requires the definition of true *multilinear models*.
A fundamental reference on the application of multilinear/tensor algebra to computer vision is (Vasilescu & Terzopoulos, 2002). The problem of disentangling the different (covariate) factors in image ensembles was there solved through the tensor extension of conventional singular value decomposition, or N-mode SVD (De Lathauwer et al., 2000).
Let us recall the basic notions of tensor algebra and multilinear SVD.
A generalization of the product of two matrices is the product of a tensor and a matrix. The *mode-n product* of a tensor $A \in \mathbb{R}^{I_1 \times \ldots \times I_n \times \ldots \times I_N}$ by a matrix $\boldsymbol{M} \in \mathbb{R}^{J_n \times I_n}$, denoted by $A \times_n \boldsymbol{M}$, is a tensor $B \in \mathbb{R}^{I_1 \times \ldots \times I_{n-1} \times J_n \times I_{n+1} \times \ldots \times I_N}$ whose entries are

$$(A \times_n \boldsymbol{M})_{i_1 \ldots i_{n-1} j_n i_{n+1} \ldots i_N} = \sum_{i_n} a_{i_1 \ldots i_{n-1} i_n i_{n+1} \ldots i_N} m_{j_n i_n}. \quad (11)$$

The mode-*n* product can be expressed in tensor notation as $B = A \times_n \boldsymbol{M}$.
A matrix is a special case of tensor with two associated vector spaces, a row space and a column space. SVD orthogonalizes these two spaces and decomposes the matrix as $\boldsymbol{D} = \boldsymbol{U}_1 \boldsymbol{\Sigma} \boldsymbol{U}^T_2$, the product of an orthogonal column space associated with the left matrix $\boldsymbol{U}_1 \in \mathbb{R}^{I_1 \times J_1}$, a diagonal singular value matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{J_1 \times J_2}$, and an orthogonal row space represented by the right matrix $\boldsymbol{U}_2 \in \mathbb{R}^{I_2 \times J_2}$.
In terms of the n-mode product, the SVD decomposition can be written as $\boldsymbol{D} = \boldsymbol{\Sigma} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2$.
"*N*-mode SVD" (De Lathauwer et al., 2000) is an extension of SVD that orthogonalizes the *N* spaces associated with an order *N* tensor, and expresses the tensor as the mode-*n* product of *N*-orthogonal spaces

$$D = Z \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \ \ldots \times_n \boldsymbol{U}_n \ \ldots \times_N \boldsymbol{U}_N. \quad (12)$$

Tensor *Z*, known as the *core tensor*, is analogous to the diagonal singular value matrix in conventional matrix SVD, but is in general a full tensor (Kolda, 2001). The core tensor governs the interaction between the *mode matrices* $\boldsymbol{U}_n$, for $n = 1, \ldots, N$. Mode matrix $\mathbf{U}_n$ contains the orthonormal vectors spanning the column space of the matrix $\boldsymbol{D}(n)$ resulting from the mode-*n* flattening of *D* (Vasilescu & Terzopoulos, 2002).
The *N*-mode SVD algorithm for decomposing *D* reads then as follows:
1. For $n = 1, \ldots, N$, compute the matrix $\boldsymbol{U}_n$ in (5) by calculating the SVD of the flattened matrix $\boldsymbol{D}(n)$ and setting $\boldsymbol{U}_n$ to be the left matrix of this SVD.
2. Solve for the core tensor as

$$Z = D \times_1 \boldsymbol{U}^T_1 \times_2 \boldsymbol{U}^T_2 \ \ldots \times_n \boldsymbol{U}^T_n \ \ldots \times_N \boldsymbol{U}^T_N. \quad (13)$$

The method was applied by Vasilescu and Terzopoulos to separate expression, pose, and identity in sets of facial images (*Tensorfaces*). They used a portion of the Weizmann face database of 28 male subjects photographed in 5 different poses under 3 illuminations performing 3 different expressions. Using a global rigid optical flow algorithm they aligned the original 512×352 pixel images to one reference image. The images were then decimated and cropped, yielding a total of 7943 pixels per image. The resulting facial image data tensor *D* was a 28×5×3×3×7943 tensor, with $N = 5$ modes.
This approach was later extended to Independent Component Analysis in (Vasilescu & Terzopoulos, 2005), where the statistically independent components of multiple linear factors were learnt.

Wang & Ahuja (2003) also made use of this technique (often called Higher-Order Singular Value Decomposition or HOSVD) for facial expression decomposition. Only three factors were considered. The crucial difference with (Vasilescu & Terzopoulos, 2002) is the suggestion to alleviate the computational load by first applying PCA to image pixel to reduce the dimensionality of the problem, leaving HOSVD to deal with the resulting principal dimensions. Recognition is implemented by measuring the cosine distance between new and learnt person or expression vectors in the respective subspaces.
Park and Savvides (2006), on their side, claimed that the use of higher-order tensors to describe multiple factors is problematic. On one side, it is difficult to decompose the multiple factors of a test image. On the other, it is hard to construct reliable multilinear models with

more than two factors as in (12). They proposed then a novel tensor factorization method based on a least square problem, and solved it using numerical optimization techniques without any knowledge or assumption on the test images. They obtained fairly good results for trilinear models.

A third alternative to multilinear modeling is a novel algorithm for positive tensor factorization proposed in (Welling & Weber, 2001). Starting from the observation that eigenvectors produced by PCA can be interpreted as modes to be linearly combined to get the data, they propose to drop the orthogonality constraint in the associated linear factorization, and simply minimize the reconstruction error under positivity constraint.
The algorithm then factorizes a tensor D of order N into F (not necessarily equal to N) *positive* components as follows

$$D_{i_1,\ldots,i_N} = \sum_{a=1}^{F} A^{(1)}_{i_1,a} \ldots A^{(N)}_{i_N,a} \tag{14}$$

so that the reconstruction error

$$\sum_{i_1,\ldots,i_N} \left(D_{i_1,\ldots,i_N} - \sum_{a=1}^{F} A^{(1)}_{i_1,a} \ldots A^{(N)}_{i_N,a}\right)^2 \tag{15}$$

is minimized. Experiments seem to show that factors produced by PTF are easier to interpret than those produced by algorithms based on singular value decomposition.
An interesting application of multilinear modeling of 3D meshes to face animation transfer can be found in (Vlasic et al., 2005). The application of multilinear algebra to the gait ID problem, though, has been pioneered by Lee and Elgammal (2005) but has not received wide attention later on. Given walking sequences captured from multiple views for multiple people, they fit a multilinear generative model using higher-order singular value decomposition which would decompose view factors, body configuration factors, and gait-style factors.
In the near future the application of positive tensor factorization or multi-linear SVD to tensorial observations like walking gaits will help the field of gait recognition progress towards a reduction of the influence of covariate factors. This will likely open the way for a wider application of gait biometrics in real-world scenarios.

Conclusions

Gait recognition is an interesting biometrics which does not undergo the limitations of other standard methods like iris or face recognition, as it can be applied at a distance to non-cooperative users. However, its practical use is heavily limited by the presence of multiple covariate factors which make identification problematic in real-world scenarios.
In this chapter, motivated by the view-invariance issue in the gait ID problem, we addressed the problem of classifying walking gaits affected by different covariates (or, equivalently, possessing different labels). We illustrated a three-layer model in which hidden Markov models with a fixed number of states are used to cluster each sequence into a fixed number of poses in order to generate the observation data for an asymmetric bilinear model. We used the CMU Mobo database (Gross & Shi, 2001) to set up an experimental comparison between the bilinear approach and other standard algorithms in view-invariant and action-invariant gait ID. We showed how bilinear modelling can improve recognition performances when the test motion is performed in an unknown style.

Natural extensions of the proposed methodology are, firstly, the representation of gait sequences or cycles as 3D tensors instead of stacked vectors. In second order the application of nonnegative tensor factorization or multidimensional SVD to gait data, in order to make identity recognition robust to the many covariate factors. This will encourage a more extensive adoption of gait identification side by side with other classical biometrics.

References

Abdelkader, C. B., Cutler, R., Nanda, H., & Davis, L. (2001). Eigengait: motion-based recognition using image self-similarity. In *Lecture Notes in Computer Science: Vol. 2091* (pp. 284–294). Berlin: Springer.

Bhanu, B., & Han, J. (2002). Individual recognition by kinematic-based gait analysis. In *Proceedings of ICPR02: Vol.* 3 (pp. 343–346).

Bouchrika, I., & Nixon, M. (2008). Exploratory factor analysis of gait recognition. In *Proc. of the 8th IEEE International Conference on Automatic Face and Gesture Recognition.*

Boutsidis, C., Gallopoulos, E., Zhang, P., & Plemmons, R.J. (2006). PALSIR: A new approach to nonnegative tensor factorization. In *Proc. of the 2nd Workshop on Algorithms for Modern Massive Datasets (MMDS).*

Cichocki, A., Zdunek, R., Plemmons, R., & Amari, S. (2007). Novel multi-layer nonnegative tensor factorization with sparsity constraints. In *Lecture Notes in Computer Science: Vol. 4432* (pp. 271–280).

Cunado, D., Nash, J.M., Nixon, M.S., & Carter, J.N. (1999). Gait extraction and description by evidence-gathering. In *Proceedings of AVBPA99* (pp. 43–48).

Cutting, J., & Kozlowski, L. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bull. Psychon. Soc.*, 9, 353–356.

Das, S.R., Wilson, R.C., Lazarewicz, M.T., & Finkel, L.H. (2006). Two-stage PCA extracts spatiotemporal features for gait recognition. *Journal of Multimedia*, 1(5), 9–17.

De Lathauwer, L., De Moor, B., & Vandewalle, J. (2000). A Multilinear Singular Value Decomposition. *SIAM Journal of Matrix Analysis and Applications*, 21(4).

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39(1), 1–38.

Elgammal, A., & Lee, C.S. (2004). Separating style and content on a nonlinear manifold. In *Proceedings of of IEEE Conference on Computer Vision and Pattern Recognition: Vol. 1* (pp. 478–485).

Elliot, R., Aggoun, L., & Moore, J. (1995). *Hidden Markov models: estimation and control*. Springer Verlag.

Ekinci, M., Aykut, M., & Gedikli, E. (2007). Gait recognition by applying multiple projections and kernel PCA. In *Proceedings of MLDM 2007*, *Lecture Notes in Artificial Intelligence: Vol. 4571* (pp. 727–741).

Frosini, P. (1991). Measuring shape by size functions. In *Proceedings of SPIE on Intelligent Robotic Systems: Vol. 1607* (pp. 122–133).

Gafurov, D. (2007). A survey of biometric gait recognition: Approaches, security and challeges. In *Proceedings of NIK-2007*.

Gross, R., & Shi, J. (2001). *The CMU motion of body (Mobo) database* (Tech. Report). Pittsburgh, Pennsylvania: Carnegie Mellon University.

Han, J., & Bhanu, B. (2004). Statistical feature fusion for gait-based human recognition. In *Proceedings of CVPR'04: Vol. 2* (pp. 842–847).

Han, J., & Bhanu, B. (2005). Performance prediction for individual recognition by gait. *Pattern Recognition Letters,* 26(5), 615–624.

Han, J., Bhanu, B., & Roy-Chowdhury, A.K. (2005). Study on View-Insensitive Gait Recognition. In *Proceedings of ICIP'05: Vol. 3* (pp. 297–300).
He, Q., & Debrunner, C. (2000). Individual recognition from periodic activity using hidden Markov models. In *IEEE Workshop on Human Motion* (pp. 47–52).
He, X., Cai, D., & Niyogi, P. (2005). Tensor subspace analysis. In *Advances in Neural Information Processing Systems 18 (NIPS)*.
Honggui, L., & Xingguo, L. (2004). Gait analysis using LLE. *Proceedings of ICSP'04*.
Huang, X., & Boulgouris, N.V. (2008). Human gait recognition based on multiview gait sequences. *EURASIP Journal on Advances in Signal Processing,* 2008.
Jafri, R., & Arabnia, H.R. (2008). Fusion of face and gait for automatic human recognition. In *Proc. of the Fifth International Conference on Information Technology.*
Johnson, A.Y., & Bobick, A.F. (2001). A multi-view method for gait recognition using static body parameters. In *Proceedings of AVBPA'01* (pp. 301–311).
Kale, A., Rajagopalan, A.N., Cuntoor, N., & Kruger, V. (2002). Gait-based recognition of humans using continuous HMMs. In *Proceedings of AFGR'02* (pp. 321–326).
Kale, A., Roy-Chowdhury, A.K., & Chellappa, R. (2003). Towards a view invariant gait recognition algorithm. In *Proceedings of AVSBS03* (pp. 143–150).
Kale, A., Sunsaresan, A., Rajagopalan, A.N., Cuntoor, N.P., Roy-Chowdhury, A.K., Kruger, V., & Chellappa, R. (2004). Identification of humans using gait. *IEEE Trans. PAMI*, 13(9), 1163–1173.
Kaziska, D., Srivastava, A. (2006). Cyclostationary processes on shape spaces for gait-based recognition. In *Proceedings of ECCV'06: Vol. 2* (pp. 442–453).
Kiers, H.A.L. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics*, 14(3), 105–122.
Kolda, T.G. (2001). Orthogonal tensor decompositions. *SIAM Journal on Matrix Analysis and Applications*, 23(1), 243–255.
Kullback, S., & Leibler, R.A. (1951). On information and sufficiency. *Annals of Math. Stat.*, 22, 79–86.
Lee, C.-S., & Elgammal, A. (2004). Gait style and gait content: bilinear models for gait recognition using gait re-sampling. In *Proceedings of AFGR'04* (pp. 147–152).
Lee, C.-S., & Elgammal, A. (2005). Towards scalable view-invariant gait recognition: Multilinear analysis for gait. In *Lecture Notes on Computer Science: Vol. 3546* (pp. 395–405).
Lee, L., & Grimson, W. (2002). Gait analysis for recognition and classification. In *Proceedings of AFGR'02* (pp. 155–162).
Lee, H., Kim, Y.-D., Cichocki, A., & Choi, S. (2007). Nonnegative tensor factorization for continuous EEG classifcation. *International Journal of Neural Systems*, 17(4), 305–317.
Li, X.L., Maybank, S.J., Yan, S.J., Tao, D.C., & Xu, D.J. (2008). Gait components and their application to gender recognition. *IEEE Trans. SMC-C,* 38(2), 145–155.
Little, J., & Boyd, J. (1998). Recognising people by their gait: the shape of motion. *IJCV,* 14(6), 83–105.
Liu, Z.Y., Sarkar, S. (2006). Improved gait recognition by gait dynamics normalization. *IEEE Trans. PAMI,* 28(6), 863–876.
Lu, H., Plataniotis, K.N., & Venetsanopoulos, A.N. (2006). Multilinear principal component analysis of tensor objects for recognition. In *Proc. of the 18th International Conference on Pattern Recognition (ICPR'06): Vol. 2* (pp. 776–779).
Lu, J.W., Zhang, E. (2007). Gait recognition for human identification based on ICA and fuzzy SVM through multiple views fusion. *Pattern Recognition Letters,* 28(16), 2401–2411.
Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T., & Yagi, Y. (2006). Gait recognition using a view transformation model in the frequency domain. In *Proceedings of ECCV: Vol. 3* (pp. 151–163).

Morup, M., Hansen, L.K., Herrmann, C.S., Parnas, J., & Arnfred, S.M. (2006). Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG. *NeuroImage*, 29(3), 938–947.
Murase, H., & Sakai, R. (1996). Moving object recognition in eigenspace representation: gait analysis and lip reading. *Pattern Recognition Lett.*, 17(2), 155–162.
Niyogi, S., & Adelson, E. (1994). Analyzing and recognizing walking figures in XYT. In *Proceedings of CVPR'94* (pp. 469–474).
Nixon, M.S., & Carter, J.N. (2006). Automatic recognition by gait. In *Proceedings of IEEE,* 94(11), 2013–2024.
Park, S.W., & Savvides, M. (2006). Estimating mixing factors simultaneously in multilinear tensor decomposition for robust face recognition and synthesis. In *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06).*
Porteus, I., Bart, E., & Welling, M. (2008). Multi-HDP: A nonparametric Bayesian model for tensor factorization. In *Proc. of AAAI 2008* (pp. 1487–1490).
Rogez, G., Guerrero, J.J., Martinez del Rincon, J., Orrite-Uranela, C. (2006). Viewpoint independent human motion analysis in man-made environments. In *Proceedings of BMVC'06.*
Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P., & Bowyer, K.W. (2005). The humanID gait challenge problem: Datasets, performance, and analysis. *IEEE Trans. PAMI*, 27(2), 162–177.
Shakhnarovich, G., Lee, L., & Darrell, T. (2001). Integrated face and gait recognition from multiple views. In *Proceedings of CVPR'01* (pp. 439–446).
Shashua, A., & Hazan, T. (2005). Non-negative tensor factorization with applications to statistics and computer vision. In *Proceedings of the 22nd International Conference on Machine Learning* (pp. 792–799).
Spencer, N.M., & Carter, J.N. (2002). Viewpoint invariance in automatic gait recognition. *Proc. of AutoID* (pp. 1–6).
Sundaresan, A., Roy-Chowdhury, A.K., & Chellappa, R. (2003). A hidden Markov model based framework for recognition of humans from gait sequences. In *Proceedings of ICIP'03: Vol. 2* (pp. 93–96).
Tan, D.L., Huang, K.Q., Yu, S.Q., Tan, T.N. (2007). Orthogonal diagonal projections for gait recognition. In *Proceedings of ICIP'07: Vol. 1* (pp. 337–340).
Tao, D. (2006). Discriminative Linear and Multilinear Subspace Methods. Ph.D. Thesis, University of London Birkbeck.
Tao, D., Li, X., Wu, X., Maybank, S.J. (2007). General tensor discriminant analysis and Gabor features for gait recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(10), 1700–1715.
Tenenbaum, J.B., & Freeman, W.T. (2000). Separating style and content with bilinear models. *Neural Computation*, 12(6), 1247–1283.
Tolliver, D., & Collins, R. (2003). Gait shape estimation for identification. In *Proc. of AVBPA'03* (pp. 734–742).
Urtasun, R., & Fua, P. (2004). *3D tracking for gait characterization and recognition* (Tech. Rep. No. IC/2004/04). Lausanne, Switzerland: Swiss Federal Institute of Technology.
Vasilescu, M.A.O., & Terzopoulos, D. (2002). Multilinear analysis of image ensembles: TensorFaces. In *Proc. of the European Conf. on Computer Vision ECCV '02* (pp. 447–460).
Vasilescu, M.A.O., & Terzopoulos, D. (2005). Multilinear independent component analysis. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05): Vol. 1* (pp. 547–553).
Veres, G., Nixon, M., & Carter, J. (2005). Modelling the time-variant covariates for gait recognition. In *Proceedings of AVBPA2005, Lecture Notes in Computer Science: Vol. 3546* (pp. 597–606).

Vlasic, D., Brand, M., Pfister, H., & Popovic, J. (2005). Face transfer with multilinear models. (Tech. Rep. No. TR2005-048). Cambridge, Massachussets: Mitsubishi Electric Research Laboratory.
Wang, L. (2006). Abnormal walking gait analysis using silhouette-masked flow histograms. In *Proceedings of ICPR'06: Vol. 3* (pp. 473–476).
Wang, H., & Ahuja, N. (2003). Facial expression decomposition. *Proceedings of ICCV* (pp. 958–965).
Welling, M., & Weber, M. (2001). Positive tensor factorization. *Pattern Recognition Letters*, 22(12), 1255–1261.
Xu, D., Yan, S., Tao, D., Zhang, L., Li, X., & Zhang, H.-J. (2006). Human gait recognition with matrix representation. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(7), 896–903.
Xu, D., Yan, S., Tao, D., Lin, S., Zhang, H.-J. (2007). Marginal Fisher analysis and its variants for human gait recognition and content-based image retrieval. *IEEE Transactions on Image Processing,* 16(11), 2811–2821.
Yam, C., Nixon, M., & Carter, J. (2004). Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, 37(5), 1057–1072.
Zhao, G., Liu, G., Li, H., & Pietikäinen, M. (2006). 3D gait recognition using multiple cameras. In *Proceedings of the 7th IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 529–534).
Zhou, X.L., & Bhanu, B. (2007). Integrating face and gait for human recognition at a distance in video. *IEEE Trans. SMC-B*, 37(5), 1119–1137.