Gait Tracking and Recognition Using Person-Dependent Dynamic Shape Model

Chan-Su Lee and Ahmed Elgammal
Rutgers University
Piscataway, NJ 08854 USA
Computer Science Department
{chansu,elgammal}@cs.rutgers.edu

Abstract

Characteristics of the 2D shape deformation in human motion contain rich information for human identification and pose estimation. In this paper, we introduce a framework for simultaneous gait tracking and recognition using person-dependent global shape deformation model. Person-dependent global shape deformations are modeled using a nonlinear generative model with kinematic manifold embedding and kernel mapping. The kinematic manifold is used as a common representation of body pose dynamics in different people in a low dimensional space. Shape style as well as geometric transformation and body pose are estimated within a Bayesian framework using the generative model of global shape deformation. Experimental results show person-dependent synthesis of global shape deformation, gait recognition from extracted silhouettes using style parameters, and simultaneous gait tracking and recognition from image edges.

1 Introduction

Characteristics of the shape deformation in a person motion contain rich information such as body configuration, person identity, gender information, and even emotional states of the person. Gait recognition has become attractive for surveillance and for security in public areas [4, 8, 2] as it is easily observable and difficult to disguise than other biometrics.

Gait involves spatiotemporal deformations in shape and appearance. Such spatiotemporal shape deformation are investigated in many appearance-based gait recognition systems [2, 14, 20, 10, 13, 22, 11]. Most of the gait recognition systems rely on silhouettes extracted using background subtraction algorithms. Recognition performance in many systems depends on the accuracy of extracted silhouettes. On the other hand, there have been a lot of work on contour tracking from cluttered environment, without the need for background subtraction, such as active shape models (ASM) [3], active contours [9], and exemplar-based tracking [21]. Spatiotemporal models are also used for contour tracking [1] However, it is difficult to achieve tracking of dynamic contour that is accurate enough to distinguish individual differences from articulated human motion. There are no spatiotemporal models for contour tracking to describe person-specific variations of shape for gait recognition.

Our objective is to model person-specific differences of shape deformation in addition to the global deformation of the shape in order to achieve adaptive tracking and person identification from gait. For certain classes of motion like gait and facial expressions, the global shape deformation might lie on a low dimensional manifold, if we consider a single person. In [7], a framework to separate the motion from the style in a generative fashion was introduced where the motion is represented in a low dimensional nonlinear manifold. Individual differences in the shape deformation can be discovered in the nonlinear mapping space between embedded representation of the configuration space and the observation. Nonlinear manifold learning can be used to find intrinsic body configuration space [23, 7]. However, when applied to image sequences, nonlinear manifold learning yields manifolds that are twisted differently according to person style, view, and other factors like clothes [6]. It is hard to achieve unified representation from these variant manifolds. In addition, the individual difference in shape contour will exist not only in the nonlinear mapping space but also in the embedding space (i.e., separation is not optimal). We propose to use kinematics manifold embedding, which represents body configuration in low dimensional space and invariant to different people, to model dynamics of shape deformation in body configuration space. The entire intrinsic configuration can have one-to-one correspondence with kinematics manifolds. Using this kinematic manifold embedding, individual difference of shape deformation can be solely contained in the nonlinear mapping. Even though the dynamics of global shape deformation are complicated and nonlinear, we can successfully model the dynamics by simple one dimensional linear model using this kinematics manifold.

We develop a generative model for person-dependent dynamic shape contour for tracking and recognition using the kinematics manifold. The generative model is represented by a configuration state and a shape style state. The shape style state is a compact representation of variations in shape contours independent of body pose (the configuration state). We use the estimated style for gait recognition. When the extracted silhouette is provided (e.g. using background subtraction), we can directly estimate the contour style state and recognize gait based on the estimated contour style parameters. On the other hand, if silhouette extraction is not possible, we use contour tracking where the tracking problem is formulated as estimation of body configuration state as well as contour style state using Bayesian framework utilizing the generative model. We can recognize person gait during tracking using particle filtering as we estimate state of *person style*, global shape deformation characteristics, as well as body configuration, which is impossible in particle model using kinematic models without global shape deformations [5]. Style estimation gradually gets discriminative using deterministic annealing like procedure in order to estimate contour style state, which can be high dimensional, robustly without trapping to local minima. Experimental results using University of Southampton gait database [18] shows potential for simultaneous gait recognition and contour tracking.

2 Decomposable Generative Models using Kinematics Manifold Embedding

We can think of the shape of a dynamic object as instances driven from a generative model. Let $z_t \in \mathbb{R}^d$ be the shape of the object at time instance t represented as a point in a d-dimensional space. This instance of the shape is driven from a model in the form

$$z_t = T_{\alpha_t}\gamma(b_t; s_t), \tag{1}$$

where the $\gamma(\cdot)$ is a nonlinear mapping function that maps from a representation of the body configuration b_t into the observation space given a mapping parameter s_t that characterizes the person shape in a way independent from the configuration and specific to the person being tracked. T_{α_t} represents a geometric transformation on the shape instance. Given this generative model, we can fully describe observation instance z_t by state parameters α_t,b_t, and s_t. Figure 1 shows a graphical model illustrating the relation between these variables where y_t is a contour instance generated from model given body configuration b_t and shape style s_t. The observed shape contour z_t is formed by geometrically transforming the contour instance y_t in the image

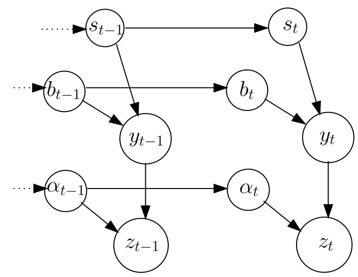

Figure 1. A graphical model for decomposable generative models

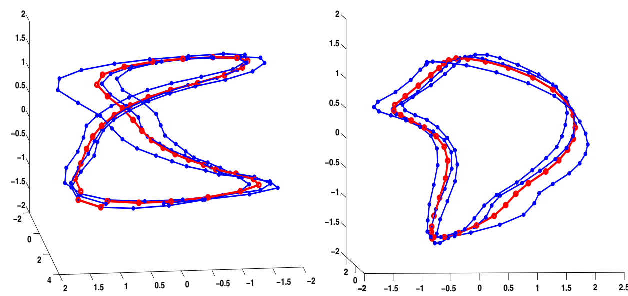

Figure 2. Kinematics manifold embedding and its mean manifold: two different views in 3D space

space. Kinematics manifold embedding is used for intrinsic manifold representation of configuration b_t.

2.1 Kinematics Manifold Embedding

We find low dimensional representation of kinematics manifold by applying nonlinear dimensionality reduction techniques for motion capture data. We first convert joint angles of motion capture data into joint locations in 3 dimensional spaces. We aligned global transformation in advance in order to count motion only due to body configuration change. In order to find low dimensional intrinsic representation from the high dimensional data (collection of joint location) we applied nonlinear dimensionality reduction procedure like Locally linear embedding (LLE) [16]. Figure 2 shows kinematics manifold based on three walking cycles of motion capture data and their mean manifold representation. The manifold is one-dimensional twisted circular manifold in three-dimensional spaces.

The manifold is represented using a one-dimensional parameter by spline fitting. For one-dimensional representation of the multiple cycles, we use mean-manifold representation in the parameterization. The mean-manifold is parameterized by spline fitting by a one-dimensional parameter $\beta_t \in \mathbb{R}$ and a spline fitting function $f : \mathbb{R} \to \mathbb{R}^3$ that satisfies $b_t = f(\beta_t)$ is used to map from the parameter space into the three dimensional embedding space as shown in Fig.2.

2.2 Decomposing and Modeling Shape Style Space

Individual variations of the shape deformation can be discovered in the nonlinear mapping space between the kinematic manifold embedding and the observation for different people. We employ nonlinear mapping based on empirical kernel map [17] to capture nonlinear deformation in difference body pose. There are three steps to model individual shape deformations using nonlinear mapping. First, for a given shape deformation sequence, we detect gait cycles and embed collected shape deformation data to the intrinsic manifold. In our case, kinematic manifold are used for gait embedding in each detected cycle. As the kinematic manifold comes from constant speed walking motion captured data, we can embed the shape sequence in equally spaced points along the manifold. Second, we learn nonlinear mapping between the kinematic embedding space and shape sequences using Generalized Radial Basis Function (GRBF) [15] similar to [7]. The mapping has the form $y_t^k = \gamma_k(b_t) = C^k \cdot \psi(b_t)$, where C^k is the mapping coefficients which depend on particular person shapes, $\psi(\cdot)$ is a nonlinear mapping with N RBF kernel functions to model the manifold in the embedding space, y_t^k is person k shape at time t, and b_t is a corresponding point on the kinematic manifold. Third, given learned nonlinear mapping coefficients $C^1, C^2, \cdots, C^K$, for training people $1, \cdots, K$, the shape style parameters are decomposed by fitting an asymmetric bilinear [19] to the coefficient space:

$$c^k = As^k, \tag{2}$$

where c^k is a vector representation of matrix C^k using column stacking. As a result, we can generate contour instance y_t^k for particular person k at any body configuration b_t as

$$y_t^k = \mathscr{A} \times s^k \times \psi(b_t), \tag{3}$$

where $\mathscr{A}$ is a third order tensor, s^k is a shape style vector for person k. The style matrix, collection of style vectors, $S = [s^1 s^2 \cdots s^k]^\mathsf{T}$, is orthonormal matrix as a result of asymmetric bilinear model analysis.

For a given cycle of walking sequence, we can estimate style vector analytically. We can find mapping coefficients C^{new} for the given new sequence based on kinematics embedding and nonlinear mapping described above. Using Eq. 2, we can find style representation in closed-form for new sequence $s^{new} = A^+ c^{new}$, where c^{new} is vector representation of matrix C^{new} and A^+ is pseudo inverse of the matrix A. Using this closed form solution, we can perform gait recognition from extracted silhouette shape sequences as will be shown in Sec. 4.2.

Ultimately the style parameter s should be independent of the configuration and therefore should be time invariant and can be estimated at initialization. However, we don't know accurate shape and body configuration initially and we cannot estimate correct shape style at the beginning. Therefore, the style needs to fit to the correct person style gradually during the tracking. So, we formulated shape style as time variant factor that should be stabilized as more observations become available. The dimension of the style vector depends on the number of people (or number of cycles) used for training and can be high dimensional. We pursue lower dimensional representation and impose constraints as explained in Sec. 3.2. The overall generative model can be expressed as

$$z_t = T_{\alpha_t}\left(\mathscr{A} \times s_t \times \psi(b_t)\right). \tag{4}$$

The tracking problem using this generative model is the estimation of parameter α_t, b_t, and s_t at each new frame given observations Z^t.

3 Bayesian Tracking and Style Estimation

Given the nonlinear shape generative model introduced above, the tracking problem is an inference problem where at time t we need to infer the body configuration b_t and the shape style s_t and the geometric transformation T_{α_t} given the observation Z_t. The Bayesian tracking framework enables a recursive update of the posterior $P(X_t|Z^t)$ over the object state X_t given all observations $Z^t = Z_1, Z_2, .., Z_t$ up to time t:

$$P(X_t|Z^t) \propto P(Z_t|X_t) \int_{X_{t-1}} P(X_t|X_{t-1}) P(X_{t-1}|Z^{t-1}) \tag{5}$$

In our generative model, the state X_t is (α_t, b_t, s_t), which uniquely describes the state of the tracked object. Observation Z_t is the captured image instance at time t.

3.1 Particle Filter with Decomposable Models

The state X_t is decomposed into three sub-states α_t, b_t, s_t. These three random variables are conceptually independent since we can combine any body configuration with any shape style with any geometrical transformation to synthesize a new shape. However, they are dependent given the observation Z_t. It is hard to estimate joint posterior distribution $P(\alpha_t, b_t, s_t|Z_t)$ for its high dimensionality. The objective of the density estimation is to estimate most likely states α_t, b_t, s_t for given observations. The decomposable feature of our generative model enables us to estimate each state by a marginal density distribution $P(\alpha_t|Z^t)$, $P(b_t|Z^t)$, and $P(s_t|Z^t)$.

We approximate the marginal density of each sub-state using maximum a posteriori (MAP) of the other sub-states, i.e.,

$$P(\alpha_t|Z^t) \propto P(\alpha_t|b_t{}^*, s_t{}^*, Z^t), \; P(b_t|Z^t) \propto P(b_t|\alpha_t{}^*, s_t{}^*, Z^t),$$

where $\alpha_t{}^*$, $b_t{}^*$, and $s_t{}^*$ are maximum a posteriori estimate of each approximated marginal density.

We represent state densities using particle filters since such densities can be non-Gaussian and the observation is nonlinear. We can represent three dimensional body configuration parameters b_t as a one-dimensional parameter β_t as explained in Sec. 2.1. The shape style is also parameterized by style class weighting parameters λ_t as in Sec. 3.2. In case of global transformation, we estimate geometric transformation parameters α_t in the image space. So, using the generative model in Eq. 4, the tracking problem is to estimate α_t, λ_t, and β_t for given observations Z^t. The marginalized posterior densities for α_t, β_t, and λ_t are approximated by three particle systems.

$$\{\alpha_t^{(i)}, {}^{\alpha}\pi_t^{(i)}\}_{i=1}^{N_\alpha}, \{\beta_t^{(j)}, {}^{b}\pi_t^{(j)}\}_{j=1}^{N_b}, \{\lambda_t^{(k)}, {}^{s}\pi_t^{(k)}\}_{k=1}^{N_s}, \tag{6}$$

where N_α,N_b, and N_s are particle numbers used for each sub-states.

3.2 Style Estimation with Constraints and Annealing Procedure

There are two factors to be considered in the shape style estimation: One is high dimensionality in the style representation and the other is discriminative power for gait recognition. The decomposition of style vector from the collection of mapping coefficient using bilinear model gives style vector whose dimension is the same as the number of cycle used for the training. We need to keep the high dimensional terms to get accurate synthesis of shape deformation as shown in Fig. 3 (b). We estimate style using deterministic annealing procedure with additional constraints in the particle sampling to achieve gradually discriminative estimation of high dimensional shape style vector.

First, we represent a new shape style as a linear combination of shape style classes within convex hull of given style classes. A new style vector s^{new} is represented by linear weighting of each of the style classes s^k, $k = 1, \cdots, K$ using linear weight λ^k with constraints:

$$s^{new} = \sum_{k=1}^{K} \lambda^k s^k \;, \; \sum_{i=1}^{K} \lambda_i^{(k)} = 1, \; \lambda_i^{(k)} \geq 0^1 \text{ for all } k \tag{7}$$

where K is the number of representative style classes. In actual style estimation using particle filtering, we force the negative $\lambda_i^{(k)}$ to be zero after important-sampling with re-sampling for the style particles. We normalize again for this modified particle values according to Eq. 7. In actual style estimation, these constraints help not to diverge into unusual dynamic shape deformation.

Second, the style estimation needs to become more discriminative as tracking progresses. If we try to be discriminative from the beginning and select a specific style for gait recognition, the estimated style inclines to be trapped in local minima. Therefore, we start from the mean style, which is the style with uniform weights for all the representative shape style classes. When additional frames become available, the estimated style vector can gradually be more discriminative since weighting particles become more sensitive to observations.

To achieve this progressive discrimination, we use a deterministic annealing like procedure: estimated style weights are forced to be close to uniform weights at the beginning to avoid hard decisions about style classes and gradually become discriminative thereafter. We assume the style distribution given observation Z_t, global transformation $T_{\alpha_t^*}$ and body configuration b_t^*, can be approximated by a Gaussian distribution.

$$\begin{aligned} {}^{s}\phi_t^{(k)} &\propto P(Z_t|\alpha_t^*, b_t^*, s_t^{(k)}) \propto \exp\left(-\frac{d(Z_t, z_t^{(k)})^2}{\Sigma_t^2}\right) \\ &= \exp\left(-\frac{d(Z_t, T_{\alpha_t^*}\mathscr{A} \times s_t^{(k)} \times \psi(b_t^*))^2}{\Sigma_t^2}\right), \end{aligned}$$

where $d(\cdot)$ is distance measure, $z_t^{(k)}$ is the contour from the generative model using $\alpha_t^*, b_t^*, s_t^{(k)}$. For geometric transformation estimation, we use weighted Chamfer distance as distance measure to give more weight to upper body part. For body pose estimation, we use oriented Chamfer distance measure to be more sensitive to leg orientation. When the variance Σ_t^2 is very big ($\Sigma_t^2 >> d(Z_t, z_t)^2$), the weight ${}^{s}\phi_t^{(k)}$ will be assigned similar value regardless to $d(\cdot)$. When the variance is small ($\Sigma^2 < d(Z_t, z_t)^2$), the likelihood is sensitive to the distance value and corresponding weights in the particle update will be discriminative. To achieve annealing-like procedure, we use style class variances, which are uniform to all classes and are defined by $\Sigma^s = T_s \sigma_s^2 I + \lambda I$ respectively as time variant parameters. The parameters T_s start with large values at the first frame and are gradually reduced and in each step and a new body configuration estimate is computed.

3.3 Tracking Algorithm

We perform tracking of gait style by sequential update of the marginalized sub-densities utilizing the predicted densities of the other sub-states. These densities are

[1]In actual implementation, we allow small extrapolation in sampled particles for fast convergence of mean particle weights to the target style

updated with current observation Z_t by updating weighting values of each sub-state particle approximations given observations. We estimate global transformation α_t using predicted density $\hat{s}_t^*$, $\hat{b}_t^*$. Then body configuration b_t is estimated using estimate global transformation α_t^*, and predicted style density $\hat{s}_t^*$. Finally style s_t is estimated with given estimation α_t^*, and b_t^*. The following table summarizes the state estimation procedure using time $t-1$ estimation.

1. Importance-sampling with re-sampling at $t-1$:
For given $t-1$ state density estimation: $\{\alpha_{t-1}^{(i)}, {}^{\alpha}\pi_{t-1}^{(i)}\}_{i=1}^{N_\alpha}$, $\{\beta_{t-1}^{(j)}, {}^{b}\pi_{t-1}^{(j)}\}_{j=1}^{N_b}$, $\{\lambda_{t-1}^{(k)}, {}^{s}\pi_{t-1}^{(k)}\}_{k=1}^{N_s}$.

Re-sampling: $\{\grave{\alpha}_{t-1}^{(i)}, 1/N_\alpha\}$, $\{\grave{\beta}_{t-1}^{(j)}, 1/N_b\}$, and $\{\grave{\lambda}_{t-1}^{(k)}, 1/N_s\}$.

2. Predict current state densities using dynamic models:

$\alpha_t^{(i)} = H\grave{\alpha}_{t-1}^{(i)} + N(0, \sigma_\alpha^2)$

$\beta_t^{(j)} = \grave{\beta}_{t-1}^{(j)} + \tilde{v}_t + N(0, \sigma_b{}^2)$, $\quad b_t^{(j)} = f(\beta_t^{(j)})$

$\lambda_t^{(k)} = \grave{\lambda}_{t-1}^{(k)} + N(0, \sigma_{s_{t-1}}^2)$, $\quad \lambda_t^{(k)} = \frac{\lambda_t^{(k)}}{\sum_{i=1}^{N_s} \lambda_{i\ t}^{(k)}}$,

3. Force style particle to satisfy constraints of Eq. 7:

If $\lambda_i^{(k)} \leq 0$ then, $\lambda_i^{(k)} = 0$ for all i,k , $\quad \lambda_t^{(k)} = \frac{\lambda_t^{(k)}}{\sum_{i=1}^{N_s} \lambda_{i\ t}^{(k)}}$,.

4. Sequential update of state weights using current observation:

Global transformation α_t with $\hat{b}_t$, $\hat{s}_t$:

$P(\alpha_t^{(i)}|\hat{b}_t^*, \hat{s}_t^*, Z_t) \propto P(Z_t|\alpha_t^{(i)}, \hat{b}_t^*, \hat{s}_t^*)P(\alpha_t^{(i)})$

${}^{\alpha}\pi_t^{(i)} = P(Z_t|\alpha_t^{(i)}, \hat{b}_t^*, \hat{s}_t^*)$, $\quad {}^{\alpha}\pi_t^{(i)} = \frac{{}^{\alpha}\pi_t^{(i)}}{\sum_{j=1}^{N_\alpha} {}^{\alpha}\pi_t^{(j)}}$

Body pose b_t with α_t, $\hat{s}_t$:

$\alpha_t^* = \alpha_t^{(i^*)}$, where $i^* = \arg\max_i {}^{\alpha}\pi_t^{(i)}$

$P(b_t^{(j)}|\alpha_t^*, \hat{s}_t^*, Z_t) \propto P(Z_t|\alpha_t^*, b_t^{(j)}, \hat{s}_t^*)P(b_t^{(j)})$

${}^{b}\pi_t^{(j)} = P(Z_t|\alpha_t^*, b_t^{(j)}, \hat{s}_t^*)$, $\quad {}^{b}\pi_t^{(j)} = \frac{{}^{b}\pi_t^{(j)}}{\sum_{i=1}^{N_b} {}^{b}\pi_t^{(i)}}$

Style s_t with α_t, b_t:

$b_t^* = b_t^{(j^*)}$, where $j^* = \arg\max_j {}^{b}\pi_t^{(j)}$

$P(s_t^{(k)}|\alpha_t^*, b_t^*, Z_t) \propto P(Z_t|\alpha_t^*, b_t^*, s_t^{(k)})P(s_t^{(k)})$

${}^{s}\pi_t^{(k)} = P(Z_t|\alpha_t^*, b_t^*, s_t^{(k)})$, $\quad {}^{s}\pi_t^{(k)} = \frac{{}^{s}\pi_t^{(k)}}{\sum_{i=1}^{N_s} {}^{s}\pi_t^{(i)}}$

5. Reducing style variance:

4 Experimental Results

We demonstrate the performance of the proposed algorithms on University of Southampton (UoS) gait database [18]. The database provides well-extracted silhouette images under controlled environments for walking sequence of more than 100 people. We used provided silhouette sequences to learn our nonlinear generative model in Sec. 2. We collected 10 subjects to learn the global shape deformations dependent on individual style and embeddings. Four cycles from each person are used to learn the style variations in each person. Total 40 cycles are used to learn the generative model ($N_s = 40$) after kinematics manifold embedding.

Representation: We represent shape by an implicit function similar to [6] where the contour is the zero level of such function. From each frame, we extracted edge using Canny edge detector algorithm.

4.1 Synthesis of New Dynamic Shapes

We tested the performance of synthesis of shape deformation according to shape style vector in our nonlinear generative model by changing style parameter and its dimension. Fig. 3 (a) shows collected original sequence of three different people. When we use reduced number of style basis, we lost details of the person. However, we still be able to generate sequences showing body pose change even with one basis as shown in the first row of Fig. 3 (b). When we used corresponding person style vectors with full dimension, the new sequence preserves detail difference of individual shape deformation. Fig. 3 (c) shows linear interpolation of style vector and corresponding shape interpolation. This capability allows tracking of new person adaptively as shown in the following experiments.

4.2 Gait Recognition Using Shape Style

We tested the performance of gait recognition in two situations. First, we perform gait recognition during tracking using edge information without any background subtraction. Gait recognition is performed by selecting highest weights in estimated style weights represented by particles. We tested the gait recognition performance for indoor sequences. The indoor sequences have relatively simple background. However, when we use just edge information, it is not easy to estimate the whole shape and identify the person as it has many missed edge in the corresponding contours and additional edge inside desired contour, which causes confusion in the estimation of shape using edge-based distance transformation (DT). Fig. 4 shows change of style weights for two people gait sequences. Both of the case, the weights begin from equal weights and gradually fit to one of shape. In case of person 2, the person style get dominant quickly. In the case of person 4, the correct person style fails to find correct style when the geometric transformation misaligned contours around 30^{th} frame. Table 1 shows gait recognition results from each person sequence. We did not count style weights of the initial 10 frames as style weights are not reliable at the beginning.

Second, we tested the gait recognition when extracted silhouette sequences are given. We selected 4 cycles from 37 people and learn the generative model. In this case, the style dimension becomes 148 (37×4) dimension. We

Table 1. Gait recognition confusion matrix

Person Id	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
P1	**1(3.3%)**	0	25 (83.3%)	0	3(10%)	1(3.3%)	0	0	0	0
P2	0	**30(100%)**	0	0	0	0	0	0	0	0
P3	0	0	**30(100%)**	0	0	0	0	0	0	0
P4	0	0	0	**23(76.7%)**	0	3(10%)	0	2(6.7%)	0	2(6.7%)
P5	0	0	0	1(3.3%)	**28(93.3%)**	0	0	1(3.3%)	0	0
P6	1(3.3%)	1(3.3%)	14(46.7%)	0	0	**0**	3(10%)	0	0	0
P7	0	0	0	0	0	0	**24(80.0%)**	5(16.7%)	0	1(3.3%)
P8	0	0	1(3.3%)	0	0	0	1(3.3%)	**28(93.3%)**	0	0
P9	1(3.3%)	0	25(83.3%)	0	0	0	1(3.3%)	2(6.7%)	**0**	1(3.3%)
P10	0	0	11(36.7%)	0	0	0	1(3.3%)	3(10%)	0	**15(50%)**

(a) Original silhouettes (b) Synthesis in different style dimension (c) Style combined synthesis

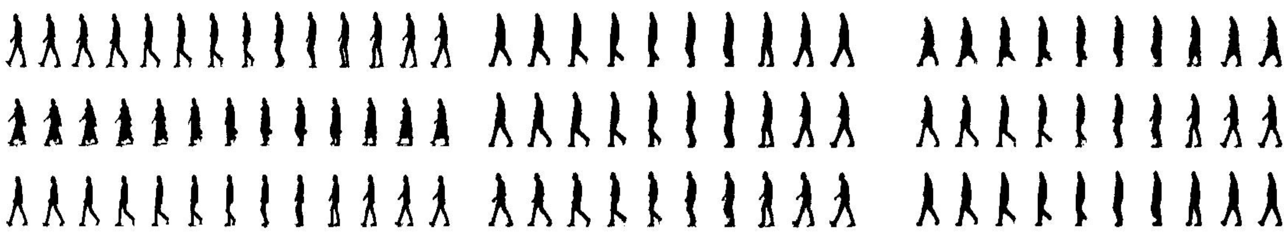

Figure 3. Style dependent dynamic shape synthesis: (a) Row 1: P1, Row 2: P2, Row 3: P3 original silhouette, (b) Synthesis of P1 silhouettes using Row 1: 1 style basis, Row 2: 25% style basis, Row 3: full style basis, (c) Synthesis by style combination: Row 1: 0.5P1+0.5P2, Row 2: 0.5P1+0.5P3, Row 3: combinations of all style vectors equally (mean style vector)

collected another 3 cycles which are not used for model learning from the same database and estimated style vector in closed form using pseudo inverse as explained in Sec. 2.2. For each estimated 148 dimensional vector, we compute similarity by inner product $S \cdot s^{est}$, which gives cosine value of two vector since the style basis are orthonomal. We classified gait by maximum similarity value and we get 83.8% recognition rate from 37 subjects by recognizing 93 sequence correctly at rank 1 among 111 (3×37) sequences. Further experiment shows cumulative matching characteristics(CMC) as in Fig. 5.

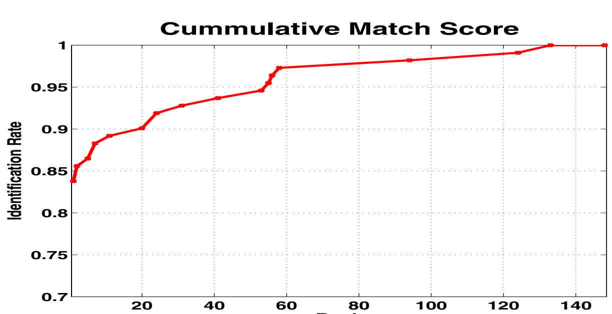

Figure 5. Performance of gait recognition using style vector

5 Conclusions

We presented a gait tracking and recognition system using nonlinear generative model for person-dependent contour deformation. Using kinematic manifold representation, we can perform tracking body pose on a one-dimensional manifold. By representing variations in spatiotemporal contour deformation among different people using style vectors, we can achieve person identification using gait simultaneously with person-adaptive contour tracking. For accurate estimation of high dimensional style vector, we added constraints in the shape style particles and employed annealing-like gradual increase of discrimination.

We performed gait recognition with simple similarity measurement and relatively small dataset that showed promising results. More advanced classification algorithms can be performed using style vectors as feature vectors. Experiments with larger date set of subject and outdoor sequences will be performed in the future. In case of view variant situation, we may be able to extend current biliear decomposition of mapping into multilinear model with view factor in addition to style factor. We may be able to achieve similar result using conceptual manifold embedding [12] instead of kinematic manifold embedding.

Acknowledgement This research is partially funded by

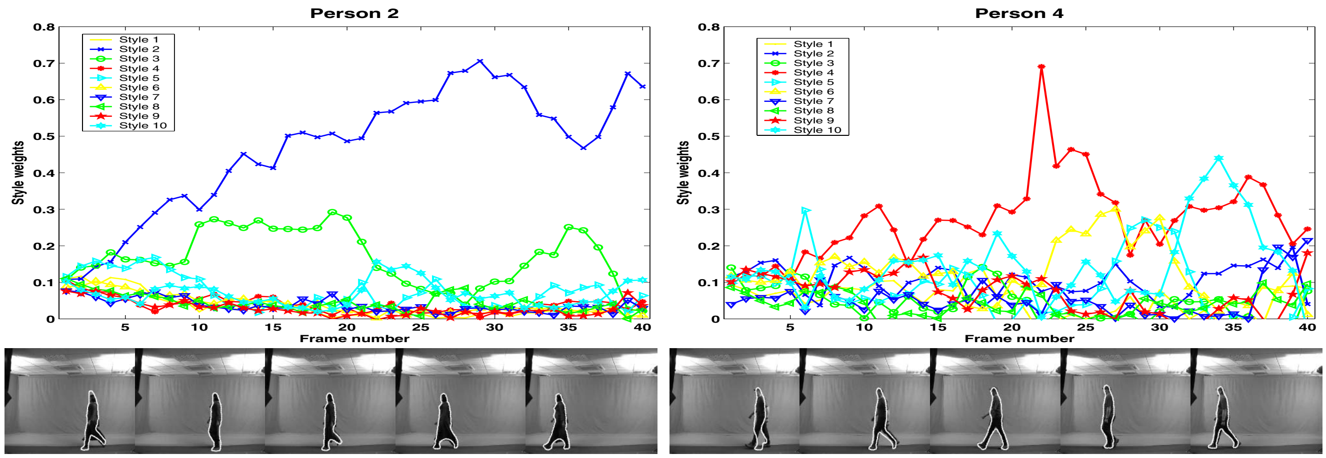

Figure 4. Simultaneous gait recognition and tracking:Left: person 2 style weights and contour tracking at 5th, 10th, 20th, 30h, and 40th frames. Right: person 4 style weights and contour tracking at 5, 10, 20, 30, 40th frame

NSF award IIS-0328991

References

[1] A. Baumberg and D. Hogg. Generating spatiotemporal models from examples. *Image and Vision Computing*, 14(8):525–532, 1996.

[2] C. BenAdbelkader, R. Cutler, and L. Davis. Motion-based recognition of people in eigengait space. In *Proc. of FGR*, pages 254–259, 2002.

[3] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models: Their training and applications. *CVIU*, 61(1):38–59, 1995.

[4] D. Cunado, M. S. Nixon, and J. Carter. Automatic extraction and description of human gait models for recognition purposes. *CVIU*, 90:1–41, 2003.

[5] J. Deutscher, A. Blake, and I. Reid. Articulated body motion capture by annealed particle filtering. In *Proc. of CVPR*, volume 2, pages 126–133, 2000.

[6] A. Elgammal and C.-S. Lee. Inferring 3d body pose from silhouettes using activity manifold learning. In *Proc. CVPR*, volume 2, pages 681–688, 2004.

[7] A. Elgammal and C.-S. Lee. Separating style and content on a nonlinear manifold. In *Proc. CVPR*, volume 1, pages 478–485, 2004.

[8] P. Huang, C. Haris, and M. Nixon. Recogising humans by gait via parametric canonical space. *Artificial Intelligence in Engineering*, 13:359–366, 1999.

[9] M. Isard and A. Blake. Condensation–conditional density propagation for visual tracking. *Int.J.Computer Vision*, 29(1):5–28, 1998.

[10] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, A. K. Roy-Chowdhury, V. Kruger, and R. Chellappa. Identification of human using gait. *IEEE Trans. Image Processing*, 13(9):1163–1173, 2004.

[11] C.-S. Lee and A. Elgammal. Gait style and gait content: Bilinear model for gait recogntion using gait re-sampling. In *Proc. of FGR*, pages 147–152, 2004.

[12] C.-S. Lee and A. Elgammal. Style adaptive bayesian tracking using explicit manifold learning. In *Proc. of British Machine Vision Conference*, 2005.

[13] L. Lee, G. Dalley, and K. Tieu. Learning pedestrian models for silhouette refinement. In *Proc. of ICCV*, pages 663–670, 2003.

[14] Z. Liu and S. Sarkar. Simplest representation yet for gait recognition: Averaged silhouette. In *Proc. ICPR*, pages 211–214, 2004.

[15] T. Poggio and F. Girosi. Networks for approximation and learning. *Proceedings of the IEEE*, 78(9):1481–1497, 1990.

[16] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linar embedding. *Science*, 290(5500):2323–2326, 2000.

[17] B. Schlkopf and A. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press, 2002.

[18] J. D. Shutler, M. G. Grant, M. S. Nixon, and J. N. Carter. On a large sequence-based human gait database. In *Proc. Int. Conf. on Recent Advances in Soft Computing*, pages 66–71, 2002.

[19] J. B. Tenenbaum and W. T. Freeman. Separating style and content with biliear models. *Neural Computation*, 12:1247–1283, 2000.

[20] D. Tolliver and R. T. Collins. Gait shape estimation for identification. In *Proc. AVBPA*, pages 734–742, 2003.

[21] K. Toyama and A. Blake. Probabilistic tracking in a metric space. In *ICCV*, pages 50–59, 2001.

[22] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Trans. PAMI*, 25(12):1505–1518, 2003.

[23] Q. Wang, G. Xu, and H. Ai. Learning object intrinsic structure for robust visual tracking. In *CVPR*, volume 2, pages 227–233, 2003.