Human Identification using Silhouette Gait Data

Chan-Su Lee CHANSU@CAIP.RUTGERS.EDU
Rutgers University, 110 Frelinghyusen Road Piscataway, NJ 08854 USA

Abstract

Identificaiton of people using a set of gait image sequences in varing situations is a challenging problem. We use a bilinear model to separate two independent factors, gait style and phase. *N-normalized gait poses* is defined and generated by embedding gait image sequences to a standard lower dimensional manifold and learning mapping from the manifold to every pixel. This normalized gait phase is used to collect aligned gait poses from different speed walking image sequence. We identify gait style-vectors, which represent factors invariant to gait pose. Gait-content vectors, dependent on the environment,are adapted to the new environment. Using a boosted gait content vector, we get a better human identification accuracy than when using the original phase vector before identifying gait content vector. Support vector machines are used to improve classification.

1. Introduction

Gait is now under investigating for biometric identification in addition to eye identification, figureprints and faces. As gait is easily observable and difficult to disguise, it has advantages compared to other biometics [1] and it becomes more attractive for surveillance and for security in public areas. As it is important to compare the performance for the same data, a baseline algorithm and data set was published [2]. The baseline algorithm computes similarities for the probe set, which is tested for new situation, and the gallery set, which is trained, silhouette sequences in disjoint subsequence frames and uses the median of the maximum correlations to represent the similarity. It shows a detection rate from 10% to 73% dependent on the covariates such as walking surface, shoe type and viewpoint. The baseline result shows that the more variation factor exist, the larger the error in human identification. We need to overcome this problem. One way is to adapt classifier or feature extractor to new environment. We can do this by adapting situation dependent factor to the new environment after learning two separate factors using bilinear model.

Gait data is sequencial and each sequence varies in walking speed and view direction. Walking speed is not constant even for an individual gait image sequences. So, we need to compare different numbers of images in variant situation to identify different people. To overcome this problem, we detect gait period and normalize the sequence for speed.

In Section 2, we cover the gait detection and normalization problems. Section 3 shows how to learn the weight matrix and two separate factor vectors. Recognition results are shown in Section 4 prior to concluding in Section 5.

2. Standarized Embedding of the Gait Cycle

This section describes how to generate aligned gait pose images from the image sequences of gait cycles. Variations in walking speed make it difficult to identify people from gait image sequences as it makes it difficult to compare gait image sequences or their features directly. A wrapping function was used [3] when comparing gait sequence by spatio-temporal correlation in order to overcome time shifting and stretching. But, it is difficult to find proper parameters if walking speed changes. In addition, we need to align gait poses among different people to find any pose-specific features common to all people. A standard embedding of the gait cycle enables us to reconstruct well-aligned gait sequences invariant to walking speed [4].

If we can learn a standardized embedding of the gait cycle, it enables us to reconstruct any intermediate gait pose. From our research related to embedding gait cycles to lower dimensional manifolds, we found that the gait manifold is one-dimensional manifold twisted in 3 dimensional space. We used a unit circle, which is a one-dimensional manifold in two dimensional space, as a standardized embedding of the

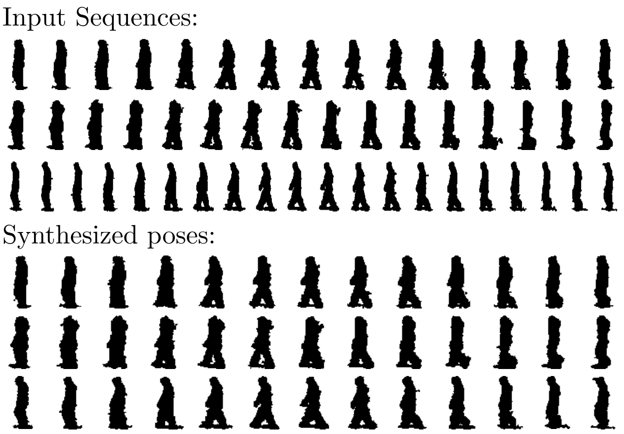


Figure 1. Original gait image sequences and their normalized gait poses

manifold. For a gait cycle with M images, we can embed each frame by M equally spaced points on the unit circle. For example, the ith frame is embedded on $\left(\cos(\frac{i\times 2\pi}{M}), \sin(\frac{i\times 2\pi}{M})\right)$ two-dimensional Cartesian coordinate point.

We learned a nonlinear mapping from the standard embedding space to the input space using generalized radial basis functions (GRBF) [5]. Nonlinear mapping from the embedding space to each individual pixel in the input space is approximated by Equation 1 that satisfies the interpolation condition $y_i^k = f^k(x_i)$ where y_i^k is the k-th pixel of imput silhouette y_i, w_i^k are real coefficients, $p^k(\cdot)$ is a linear polynomial, and $|\cdot|$ is the norm on R^2. The mapping coefficients can be obtained by solving a linear system of equations. Such mapping can be written in the form of a generative model as in Equation 2. We can generate interpolated image any given point on the embedding manifold using the generative model.

$$f^k(x) = p^k(x) + \sum_{i}^{M} w_i^k \phi(|x - x_i|) \tag{1}$$

$$f(x) = B \cdot \psi(x) \tag{2}$$

M sequential images for a gait cycle are sampled images of a single continuous gait cycle into M sampled poses with equal intervals. We define the *N-normalized gait poses* as a collection of N gait pose images from continuous one cycle gait motion by equal time intervals. If the gait cycle time is T, then it is composed of N pose images with interval of T/N. This normalized gait poses overcome variation of speed and generate aligned poses in different gait cycle times.

As a first step toward generating normalized gait poses, we need to detect the gait period T. We used variations of body width in image sequence to detect gait cycle. We used two kinds of filtering to find the proper variation of body width. First, we do median filtering for the background-subtracted silhouette images. The filter size was adjusted to remove irrevelant image features to the body width such as noise blobs and shadows. The width values are still very noisy and we smooth the width signal by convolution of one-dimensional Gaussian filter. Our experiment shows that sometimes the gait period shows $2 \sim 5$ frame variations within person in the NIST-UFL gait data. There are more variations between people.After detection of the gait period T, we standardized the gait cycle to reconstruct the normalized gait poses. In figure 1, first three rows show original image sequences for three different people and next three show 13-synthesized gait poses synthesized using the learned models from each input sequences.

3. Learning A Bilinear Model

It is well known in psychology that human perceptual systems naturally separate the content and style factor of their observations in identifying a familiar face or gait seen under unfamiliar viewing conditions. Bilinear models identify two independent factors [6]. Usually, one factor is called *style* and the other is called *content.* We call the*gait style* factor the gait characteristics in a cycle, which are invariant and allow us to identify different people independently of other conditions. The *gait content* factor is the variable factor in a gait cycle, which is dependent on the time sequence and other conditions such as viewpoint, shoe, and background.

It is necessary to collect data sets that are well aligned by two factors in order to learn separate factors. We extract L gait cycles from individual gait image sequences using an automatic gait period detection algorithm. N-normalized gait poses are generated from each gait cycle, which makes it possible to align gait cycles from different numbers of frame into fixed number of image sequences aligned by the same gait pose. If the image size of one frame is K, then the data dimension of J number of people normalized gait poses will be $K \times N \times L \times J$. The number of gait cycles is $S = L \times J$. We organized this gait-image-sequence data set into two forms: One is the *style format*, the other is the *content format.* In style format, each column contains N gait phase image sequences to one gait cycle and the column vector size is $K \times N$. Each column in phase format represents images of one gait pose from all of the different gait cycles. When we use an N normalized-gait cycle, N column vectors are collected with size $K \times S$.

3.1. Learning gait style factor

Learning the gait style factor means fiting a linear model to the style format data as in Equation 4, which means finding T_{ps} and S to minimize error $E = ||D_{sf} - T_{ps}S||^2$. The solution can be found by singular value decomposiiton (SVD). If we compute SVD of $D_{sf} = UDV^T$, the least-square optimal solution is $S = V^T$ and $T_{ps} = UD$. We can do J-dimensional approximation by choosing first J largest diagonal terms in S and making the others zero. If the training data are not equally collected across two different factors, we need to use an other numerical method like quasi-Newton optimization to find a solution to minimize error [6].

$$I_g = \sum_{i=1}^{n} \sum_{j=1}^{s} w_{ij} p_i s_j, \tag{3}$$

$$D_{sf} = T_{ps} S, \tag{4}$$

$$D_{sf} = PW_{ps}S, \; D_{pf} = SW_{sp}P. \tag{5}$$

Each style vector captures the gait style factor that is common to all the gait-pose variation in each cycle. This gait style vector can be a feature vector that is invariant to pose and situation variation. After learning the gait content vector and its transfer matrix, we can find new style vector representation for a new gait-cycle image sequence I_{new} by taking the pseudoinverse of the transfer matrix, as in Equation 6. This new style vector for a gait cycle can be used as a new feature vector to recognize gait.

$$s_{new} = D^{-1} U^T I_{new} \tag{6}$$

3.2. Learning gait style factor and gait phase factor

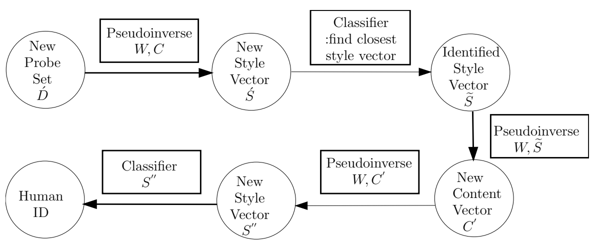


Figure 2. Classifier with adaptation of content vectors

To fit the gait style factor and the gait content factor from an image sequence data set, we need to learn the S and P vectors together. This is known as two-mode component analysis [7] in case of scalar variable, and when extended to the case of vector components using vector transpose [6], a bilinear model. The symmetric model enables us to adapt gait phase vectors to reflect changes of environment. In a style factor learning approach such as in Equation 4, it is difficult to modify whole transfer matrix T_{ps}. The symmetric model learning was done iteratively using vector transposition and SVD.

3.3. Adapting gait content factor

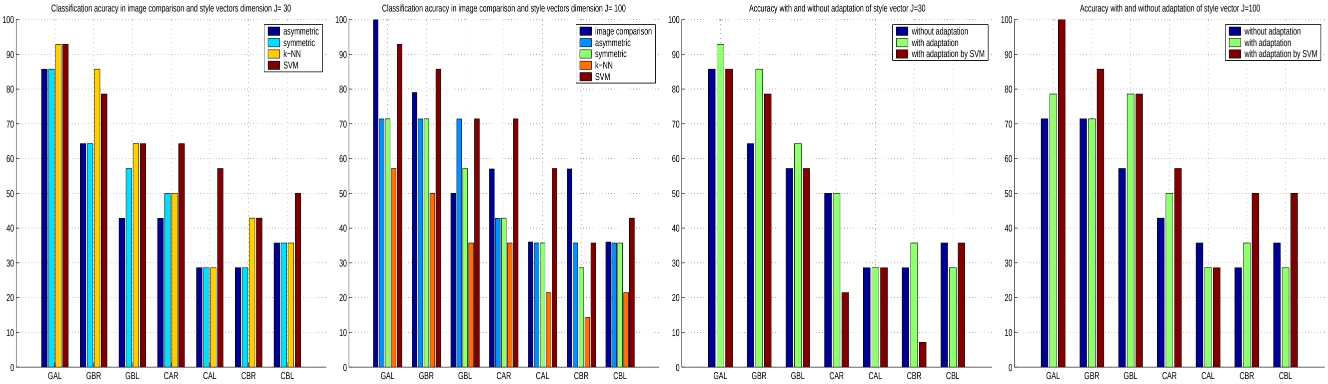
To identify the style vector for the test data set (probe data set), we do preclassification using the current symmetric model. We choose a closest gait style vector from the training set for each cycle in order to approximate original gait style vectors before learning gait content vectors. Figure 2 shows procedures to identify new content vectors and estimation of original style vector using approximated style vector.

4. Human Identification using gait style vectors

The NIST-USF Gait database [2] is used to learn a bilinear model and test our algorithm. We used computed silhouettes for the May-Nov-2001 data. This data set has variations due to viewpoint, footwear, walking surface and with/without briefcase. It also provides optimized baseline performance, which enables precise comparision with other algorithms. We selected 14 peoples as a gallery set and tested 7 different conditions by variation of viewpoint, footwear, and walking surface. Original silhouette image size is 128×88 and we resized to 64×44. In our experiment, the image vector size is $64 \times 44 = 2816$, the number of gait pose in normalized gait phase is 13, the vector size for one gait cycle is $36608 = 2816 \times 13$. The number of style vectors is $112 = 8$(cycle per people)$\times 14$ (people). The dimension of each data set D_{sf} is 36608×112.

We perform a naive identification using normalized gait data to evaluate the effectiveness of the gait normalization method. As one person gait image sequences are composed of several cycle, each person is identified by finding minimum average Euclidean distance of normalized gait image vector.Table shows the recognition result for 14 people. It shows better recognition compared with baseline result.

To identify people using a gait style vector, we find a new gait style vector for each cycle by Equation 6. We can identify people by gait style vector instead of using the image sequence itself, which is very expensive and not good for regular classifier for its high dimensionality. We tested the accuracy of identification with a k-nearest neighbor classifier (k-NN) and support vector machines (SVMs). In human identification using gait sequences, we need to determine people based on



(a) J=30 (b) J=100 (c) J=30 (d) J=100

Figure 3. a,b: classification rates based on recovered style vectors. c,d: classification rates with/without adaptation to new content vectors

the whole image sequences, composed of several cycles. We implemented a classifier to identify people in each sequence, and identified people by selecting the majority of individual classification result. Figure shows identification ratio in different number of dimension. Figure 3 (a) (b) shows improvement of SVMs over kNNs in high dimensional case (j=100) even though it is similiar in low dimensional case (J = 30).

Difference	Baseline	Normalized Gait
View	73%	100%
Shoe	78%	79%
Shoe, view	48%	50%
Surface	32%	57%
surface, view	17%	36%
Surface, shoe	22%	57%
Surface, shoe, view	17%	36%

We also tested how well the adaptation of the gait style in new environment helps classification in varying situations. Figure 3 (c)(d) graph shows a comparison of identification accuracy of symmetric model with adjusting-gait style vector.It shows improvement of recognition when we modify gait phase vector to new environment and find new gait style for classification. If we can use a better classifier in the preclassification, we can expect better adaptation to an unknow environment.

5. Conclusion and Future Work

We generate a well aligned normalized-gait image sequence by standarad unit circle manifold embedding and nonlinear mapping. This allowed us to learn two factors, a gait style factor and a gait content factor. This adaptation shows improvement of identification. In the future, we are going to test with more human data with maximum available cycles in every image sequences. In this case, we need to learn bilinear models based on EM algorithms.

References

[1] David Cunado, Mark S. Nixon, and John Carter. Automatic extraction and description of human gait models for recognition purposes. *Computer Vision and Image Understanding*, 90:1–41, 2003.

[2] P. Johathon Phillips, Sudeep Sarkar, Isidro Robledo, Patrick Grother, and Kevin Bowyer. Baseline results for the challenge problem of human id using gait analysis. In *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 137–142, 2002.

[3] Hiroshi Murase and Rie Sakai. Moving object recognition in eigenspace representation: gait analysisand lip reading. *Pattern Recognition Letters*, 17:155–162, 1996.

[4] Ahmed Elgammal and Chan-Su Lee. Separating style and content on a nonlinear manifold. In *Submitted to CVPR 2004 Conference*, 2004.

[5] Tomaso Poggio and Federico Girosi. A theory of networks for approximation and learning. Technical Report A.I. Memo No.1140, MIT Artificial Intelligence Laboratory, 1989.

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

[7] Jan R. Magnus and Heinz Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, 1988.