Simplest Representation Yet for Gait Recognition: Averaged Silhouette

Zongyi Liu and Sudeep Sarkar
Computer Science and Engineering
University of South Florida; Tampa; FL 33620
{zliu4, sarkar}@csee.usf.edu

Abstract

We present a robust representation for gait recognition that is compact, easy to construct, and affords efficient matching. Instead of a time series based representation comprising of a sequence of raw silhouette frames or of features extracted therein, as has been the practice, we simply align and average the silhouettes over one gait cycle. We then base recognition on the Euclidean distance between these averaged silhouette representations. We show, using the recently formulated gait challenge problem (www.gaitchallenge.org), that the improvement in execution time is 30 times while possessing recognition power that is comparable to the gait baseline algorithm, which is becoming the comparison standard in gait recognition. Experiments with portions of the average silhouette representation show that recognition power is not entirely derived from upper body shape, rather the dynamics of the legs also contribute equally to recognition. However, this study does raise intriguing doubts about the need for accurate shape and dynamics representations for gait recognition.

1. Introduction

Possibilities for gait recognition was demonstrated in the early 70s using light point displays. Over the past few years, a variety of approaches to gait recognition have been proposed, almost all of which are based on matching silhouettes of persons. Approaches to gait recognition, typically, involves matching time series of features extracted in each frame. The possible frame features include just the raw silhouettes as in the gait baseline algorithm [8], shape PCA coefficients [15], shape moments [12], silhouette width vector [11, 4, 13], and body part ellipses [10]. The matching of the trajectories of these features rely on simple spatio-temporal correlation [15, 8, 11], or matching maps of silhouette correlations [1], or dynamic time warping and HMM [4, 13]. Apart from these classes of approaches that tend to emphasize both the shape of the silhouette and its evolution over time, there are approaches that emphasize just the shape [3, 14] or use static body parameters [6] with almost equal or better performance than the first class of approaches.

Given this collective wisdom about gait recognition, it is pertinent to ask: what is the simplest representation that suffices for gait recognition? The quest for the simplest representation is meaningful from both a computational point of view and from a robustness point of view; simpler ideas tend to be generalizable across a wider range of conditions. Towards this end, we propose the averaged silhouette representation; we simply consider the sum of the silhouettes over approximately one gait cycle as the gait representation. The matching process simply considers the Euclidean distances between these average silhouettes. The representation is robust with respect to gait cycle length estimates and does not depend on the choice of the starting stance of the gait cycle. There is no need for stance matching or gait alignment before matching. The idea behind the proposed representation is somewhat similar to the summed symmetry maps proposed in [5], where bilateral symmetry map of each silhouette is first extracted and then summed over one gait cycle. We, however, do not even extract the symmetry maps. The use of cumulative images for motion-based human activity recognition is not new. Bobick and Davis [2] used temporal template, a vector-image constructed by weighted image-differencing through the motion history, to identify different human activities, such as sit-down, arms-wave, and crouch down. We show that this kind of representation seem to be sufficient also for recognition.

We use the HumanID Gait Challenge framework [8] to demonstrate the efficacy of the proposed representation. The challenge problem, which is being used by the gait community [13, 14, 9], consists of a baseline algorithm, a set of twelve experiments (A through L), and a large data set (1870 sequences, 122 subjects, 1.2 Terabytes of data). The baseline gait recognition algorithm estimates silhouettes by background subtraction using the Mahalanobis distance in the RGB color space and performs recognition by spatio-temporal correlation of the silhouettes. The twelve challenge experiments are of increasing difficulty and examine the effects of five covariates (and some of their combinations) on performance. The covariates are: change in viewing angle, change in shoe type, change in walking surfaces (concrete and grass), carrying or not carrying a briefcase, and temporal differences. The detailed

0 This research was supported by funds from DARPA (F49620-00-1-00388) and NSF (EIA 0130768).

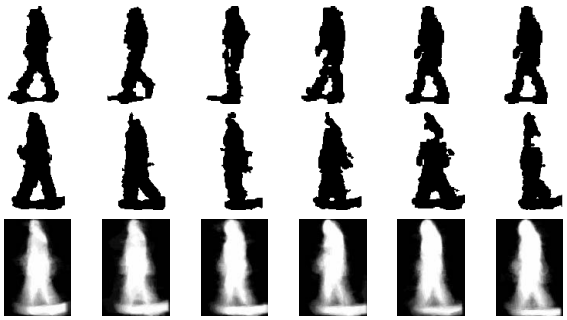

Figure 1. The first and second rows show samples of the binary silhouettes over one gait cycle, for two subjects, respectively. The third row shows the averaged silhouettes for the subject in the second row; each averaged over a different gait cycle.

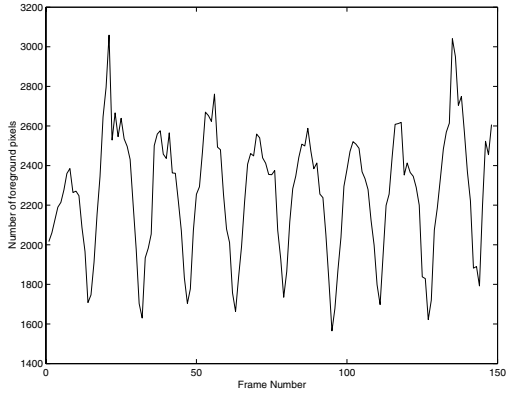

Figure 2. Variation over time of the number of foreground pixels from the bottom half of the silhouettes.

specifications of the experiments, the gait data, the source code of the baseline algorithm, and scripts to run, score, and analyze the challenge experiments were taken from `www.GaitChallenge.org`.

2. Averaged Silhouette Representation

The first step is silhouette extraction in each frame based the Mahalanobis distance from the background pixel statistics. We compute the background statistics of the RGB values at each image location, (x,y), in terms of the mean $\mu_B(x,y)$ and the covariances $\Sigma_B(x,y)$ of the RGB values at each pixel location. Using the Mahalanobis distance of a pixel value as the observation, pixels are classified into foreground or background using Expectation Maximization (EM) with a Gaussian mixture model. We found that the process stabilizes within 15 or so iterations. Fig. 1 shows some example silhouettes.

The second step is to estimate the gait periodicity, N_{gait}. We simply count the number of foreground pixels in the silhouette in each frame over time, $N_f(t)$. This number will reach a maximum when the two legs are farthest apart (full stride stance) and drop to a minimum when the legs overlap (heels together stance). To increase the sensitivity, we consider the number of foreground pixels mostly from the legs, which are selected simply by considering only the bottom half of the silhouette. Fig. 2 shows an instance of the variation of $N_f(t)$. Notice that two consecutive strides constitute a gait cycle. We compute the median of the distances between minima, skipping every other minima. Using this strategy, we get two estimates of the gait cycle, depending on whether we skipped the first minima or not. We estimate the gait period by the average of these two medians. We have observed that the estimate are pretty robust even in the presence of shadows, the size of which tend to be correlated with the stance.

The third step is average silhouette computation. Given a sequence of silhouettes, $\mathbf{S} = \{\mathbf{S}(1), \cdots, \mathbf{S}(M)\}$, we partition it into subsequences of gait period length, denoted by $\mathbf{S_{Pk}} = \{\mathbf{S}(k), \cdots, \mathbf{S}(k+N_{Gait})\}$. For each subsequence we average the silhouettes to arrive at a set of average silhouettes, $\mathbf{AS}(i), i = 1, \cdots, \lfloor \frac{M}{N_{Gait}} \rfloor$.

$$\mathbf{AS}(i) = \frac{1}{N_{Gait}} \sum_{k=iN_{Gait}}^{(i+1)N_{Gait}-1} \mathbf{S}(k) \tag{1}$$

Fig. 1 shows examples of the average silhouette representation for a sequence. Note that this representation implicitly captures the shape of the template and, to a lesser extent, the temporal dynamics of gait. The time spent at each stance shows up indirectly as intensity in the average silhouette representation.

3. Similarity Computation

For gait recognition, we need to compute the similarity between a given probe sequence and a stored gallery sequence. Let the average silhouettes from a probe and a gallery be denoted by $\{\mathbf{AS_P}(i) | i = 1, \cdots, N_P\}$ and $\{\mathbf{AS_G}(j) | j = 1, \cdots, N_G\}$, respectively. The similarity is defined as the negative of the median of the Euclidean distance between the averaged silhouettes from the probe and the gallery.

$$\mathrm{Sim}(\mathbf{AS_P}, \mathbf{AS_G}) = -\mathrm{Median}_{i=1}^{N_p} \left(\min_{j=1}^{N_G} ||\mathbf{AS_P}(i) - \mathbf{AS_G}(j)|| \right) \tag{2}$$

4. Gait Challenge Performance

We use the HumanID Gait Challenge framework [8] to demonstrate the efficacy of the proposed representation.

The challenge problem consists of a set of twelve experiments (A through L) of increasing difficulty and examine the effects of five covariates (and some of their combinations) on performance. After the widely accepted face recognition technology evaluations (FERET) [7], each experiment is specified in terms of a gallery set and a probe set consisting of data from a set of subjects, with controlled difference in covariates. The covariates are: change in viewing angle, change in shoe type, change in walking surfaces (concrete and grass), carrying or not carrying a briefcase, and temporal differences. Of these 12 experiments, we pick 5 key experiments, exercising variation in just one covariate at a time. For instance, the first experiment (A), which studies the effect of viewpoint, consists of a gallery of sequences taken from the left view and probes of sequences from the right view.

We list performance in terms of the correct identification rates at the top most rank, i.e. fraction of times the correct match to a probe is the top ranked match among all the matches of that probe to the complete gallery set. This is a standard performance metric used in biometrics [7] for the identification scenario, where one is interested to find a match to a given probe from the whole gallery set, i.e. one-to-many match. (For the verification scenario, where one is interested in matching one probe to one gallery (one-to-one match), the performance is specified in terms of standard false alarm and detection rates. In general, identification is considered to be a harder problem than verification.) We have found that the pattern of verification performance is similar to that for identification, so we do not report those here due to space limitations.

In Fig. 3, we plot CMCs for the first 20 ranks for the gait baseline algorithm, which uses spatio-temporal correlation of the silhouette, and recognition based on the averaged silhouette representation proposed here. We see that performance on three of the experiments, i.e. A (view), B (shoe), and H (carry), is better with averaged silhouettes. There is some fall in performance for the other two experiments exercising surface (D) and time (K). However, statistical McNemar's tests show that the rank 1 identification rates are not statistically significant (P-value > 0.05). On a 800 MHz SunFire server it took 4.63s on an average to compare two sequences by spatio-temporal correlation as compared to 0.14s on an average to compare similarity using the average silhouette; a 30 times improvement in time.

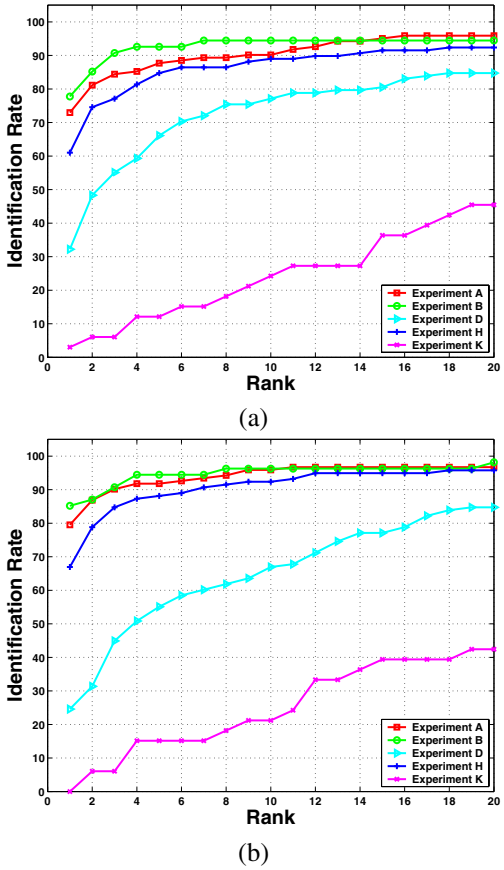

Figure 3. Performance on 5 key experiments from the gait challenge problem. Identification performance in terms of CMCs with a gallery set of 122 subjects of (a) the baseline algorithm using individual silhouette frames, (b) the averaged silhouettes.

5. Discussion and Conclusions

We have seen that even though the averaged silhouette representation is simple to compute and is compact, it has significant recognition power. We demonstrated this using the recently formulated gait challenge problem; we showed performances are statistically equivalent to the baseline algorithm for gait recognition. In this context, it is worth noting that the gait baseline performance is competitive with respect to other gait recognition algorithms that are being proposed. For example, on experiment D that compares sequences from different surfaces, the identification rate of (i) continuous HMM based recognition [13] is 36%, (ii) body shape [14] is 21%, and (iii) body part based recognition [9] is 25%, reported on a smaller version of the gait challenge problem involving just 71 subjects. The rate of the baseline algorithm on the corresponding data subset is 29% [8].

The competitive performance of the averaged silhouette representation raises intriguing questions about the importance of shape vs. dynamics in gait recognition. Are we recognizing persons from the upper body shape? This is a tough question to answer since the arm dynamics is merged with upper body shape. However, we ran a simple experiment that suggest that leg dynamics do contribute to recognition. We removed (or kept) portions from the top of the av-

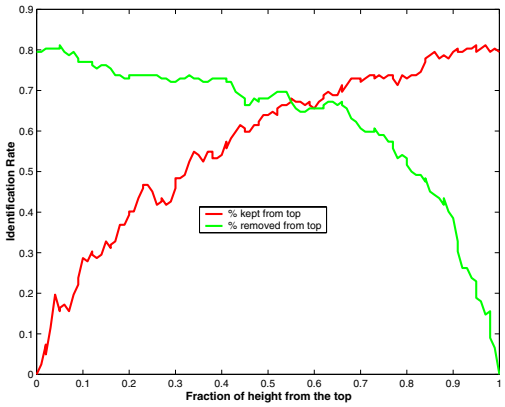

Figure 4. Variation in identification rates as different portions of the silhouettes are removed.

eraged silhouette representation. In Fig. 4, we plot the identification rate (at rank 1) for Experiment A (view change) versus the percentage of the silhouette removed (green) or kept (red) from the top. We see that the lower portion of the average silhouette, which is dominated by leg dynamics, contributes equally to recognition as the top half portion, which is dominated by body shape.

Another factor that might be confounding recognition is the presence of shadows and background subtraction errors. Are correlations in error patterns contributing to recognition? Error patterns will tend to be correlated if the sequences being compared are collected roughly around the same time or subjects do not change clothing. Maybe these error patterns are getting reinforced in the averaged representation, thus contributing to recognition. To help us answer this question, we *manually* created silhouettes over one gait cycle from 71 subjects in the gait challenge problem for experiments studying shoe variation (B), surface variation (D), carrying condition (H), and time (K). The manual silhouettes are "clean" without shadow or background subtraction artifacts. Table 1 lists the rank 1 identification rates with individual manual silhouettes and with the averaged silhouette for these four experiments. We see that performances are very close; the rank 1 performances are not statistically different, as judged by McNemar's test with a P-value of 0.05.

Exp	B (shoe)	D (surface)	H (carry)	K (time)
Seq.	49%	20%	10%	12%
Avg.	54%	14%	25%	3%

Table 1. Identification rate at rank 1 for a gallery size of 71 subjects obtained by correlating sequences of *manual* silhouette frames and using averaged *manual* silhouettes.

References

[1] C. BenAbdelkader, R. Cutler, and L. Davis. Motion-based recognition of people in eigengait space. In *International Conference on Automatic Face and Gesture Recognition*, pages 267–272, 2002.

[2] A. Bobick and J. Davis. Real time recognition of activity using temporal templates. In *IEEE Workshop on Applications of Computer Vision*. 1996.

[3] R. Collins, R. Gross, and J. Shi. Silhouette-based human identification from body shape and gait. In *International Conference on Automatic Face and Gesture Recognition*, pages 366–371, 2002.

[4] N. Cuntoor, A. Kale, and R. Chellappa. Combining multiple evidences for gait recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2003.

[5] J. Hayfron-Acquah, M. Nixon, and J. Carter. Automatic gait recognition by symmetry analysis. In *International Conference on Audio- and Video-Based Biometric Person Authentication*, pages 272–277, 2001.

[6] A. Johnson and A. Bobick. A multi-view method for gait recognition using static body parameters. In *International Conference on Audio- and Video-Based Biometric Person Authentication*, pages 301–311, 2001.

[7] P. Jonathon Phillips, H. Moon, S. Rizvi, and P. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

[8] P. Jonathon Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer. The gait identification challenge problem: Data sets and baseline algorithm. In *International Conference on Pattern Recognition*, pages 385–388, 2002.

[9] L. Lee, G. Dalley, and K. Tieu. Learning pedestrian models for silhouette refinement. In *International Conference on Computer Vision*, 2003.

[10] L. Lee and W. Grimson. Gait analysis for recognition and classification. In *International Conference on Automatic Face and Gesture Recognition*, pages 155–162, 2002.

[11] Y. Liu, R. Collins, and Y. Tsin. Gait sequence analysis using frieze patterns. In *European Conference on Computer Vision*, May 2002.

[12] J. Shutler, M. Nixon, and C. Carter. Statistical gait description via temporal moments. In *4th IEEE Southwest Symp. on Image Analysis and Int.*, pages 291–295, 2000.

[13] A. Sunderesan, A. K. Roy Chowdhury, and R. Chellappa. A hidden markov model based framework for recognition of humans from gait sequences. In *IEEE International Conference on Image Processing*, 2003.

[14] D. Tolliver and R. Collins. Gait shape estimation for identification. In *International Conference on Audio- and Video-Based Biometric Person Authentication*, 2003.

[15] L. Wang, W. Hu, and T. Tan. A new attempt to gait-based human identification. In *International Conference on Pattern Recognition*, volume 1, pages 115–118, 2002.