Evaluation of Gait Recognition

SUDEEP SARKAR¹, ZONGYI LIU²
¹Computer Science and Engineering, University of South Florida, Tampa, FL, USA
²Amazon.com, Seattle, WA, USA

Synonyms

Gait recognition; Progress in gait recognition

Definition

Gait recognition refers to automated vision methods that use video of human gait to recognize or to identify a person. Evaluation of gait recognition refers to the benchmarking of progress in the design of gait recognition algorithms on standard, common, datasets.

Introduction

Design of biometric algorithms and evaluation of performance goes hand in hand. It is important to constantly evaluate and analyze progress being at various levels of biometrics design. This evaluation can be of three types: at algorithm-level, at scenario-level, and at operational-level, roughly corresponding to the maturity of the biometric. Given the young nature of gait as a biometric source, relative to the mature biometrics such as fingerprints, current evaluations are necessarily at algorithm-level. The motivation behind algorithm-level evaluations is to explore possibilities, to understand limitations, and to push algorithmic research towards hard problems. Some of the relevant questions are

1. Is progress being made in gait recognition of humans?
2. To what extent does gait offer potential as an identifying biometric?
3. What factors affect gait recognition and to what extent?
4. What are the critical vision components affecting gait recognition from video?
5. What are the strengths and weaknesses of different gait recognition algorithms?

An overview of the current evaluation of gait as a potential biometric is discussed here, with particular emphasis on the progress with respect to the HumanID gait challenge problem that has become the de-facto benchmark. A synthesis of gait recognition performances reported on this dataset and other major ones is provided here, along with some suggestions for future evaluations.

A Panoramic View of Performance

To take stock of the progress made in gait recognition, consider a summary of the identification rates reported in the recent literature on different kinds of publicly available experimental protocols and datasets (>25 persons) such as the CMU-Mobo dataset [1] (indoor, 25 subjects), the UMD dataset [2] (outdoor, 55 subjects), the Southampton Large dataset [3] (indoor and outdoor, 115 subjects), the CASIA Gait Dataset [4] (indoor, 124 subjects), and the HumanID Gait Challenge dataset [5] (outdoor, 122 subjects). Figure 1 lists the average identification rates for matching across different conditions, i.e., the ▶ probe and the gallery differed with respect to the indicated ▶ covariate. Of course, the caveat is that the conclusions are conditioned on the kinds of variations of each covariate observed in the respective datasets. Hence, a definitive conclusion is hard to make. However, this kind of summary has some conclusive weight since, it encompasses the findings of multiple research groups. It should provide some directions for focusing future research.
data shows that outdoor gait recognition, recognition across walking surface-type change, and recognition across months are all hard problems. Clothing, footwear, carrying condition, and walking speed does not seem to be hard covariates to overcome. As expected, performance also drops with dataset size, which suggests that it is imperative to demonstrate the efficacy of an idea on as large a dataset as possible.

A deeper look at the performances reported on commonly available datasets, in particular the HumanID gait challenge problem, will form the basis for more definitive conclusions about the progress that is being made.

**The HumanID Gait Challenge Problem**

The development of gait biometrics is following a path that is somewhat different from other biometrics, for which serious evaluation benchmarks appeared after years of algorithmic development. It was more than 20 years for face recognition, whereas evaluation framework for gait recognition appeared in less than 10 years after the first publication of vision algorithms for gait recognition. Bulk of the research in gait recognition was spurred by the DARPA HumanID at a distance program. The HumanID gait challenge problem was formulated in this program to facilitate objective, quantitative measurement of gait research progress on a large dataset [5]. As of end of 2007, this dataset has been distributed to more than 40 research groups. Many gait recognition research papers report performance on this dataset.

This challenge problem does not just consist of a dataset, but also provides a well-defined experimental framework for others to follow, along with an established benchmark.

**The Dataset**

The data was collected outdoors. For each person in the data set, there are combination of as many as five conditions or covariates. The conditions are: (1) two camera angles (L and R), (2) two shoe types (A and B), (3) two surfaces (grass and concrete), (4) with and without carrying a briefcase (B or NB), and (5) two different dates 6 months apart, May and November. The covariates were chosen based on consultation with gait recognition researchers in the Human ID program.
These are, of course, not the only variables that can impact gait, but were logistically feasible and likely to impact gait the most. Attempt was made to acquire a person’s gait in all possible combinations, and there are up to 32 sequences for some persons. Hence, the full dataset can be partitioned into 32 subsets, one for each combination of the five covariates. The partitioning of the data is visualized in Fig. 2. Each cell refers to a unique combination of view, shoe type, and surface covariates. The smaller arrangement of cells represent the data from repeat subjects. Comparisons between these subsets are used to set up challenge experiments; more on this later. The full data set consists of 1,870 sequences from 122 individuals. This dataset is unique in the number of covariates exercised. It is the only data set to include walking on a grass surface. Figure 3 shows some sample frames from this dataset.

In addition to the raw data sequence, there is an ancillary information associated with the data. First, for each sequence, there is meta-data information about the subject’s age, sex, reported height, self reported weight, foot dominance, and shoe information.

Second, for a subset of this dataset, manually created silhouettes (see Fig. 4) are available. These manual silhouettes should not be used to test any recognition algorithm, but they could be used to build models or to study segmentation errors. More details about the process of creating these manual silhouettes and the quality checks performed can be found in [6]; here are some salient aspects. Seventy one subjects from one of the two collection periods (May collection) were chosen for manual silhouette specification. The sequences corresponding to these subjects were chosen from the (1) gallery set (sequences taken on grass, with shoe type A, right camera view), (2) probe B (on grass, with shoe type B, right camera view), (3) probe D (on concrete, with shoe type A, right camera view), (4) probe H (on grass, with shoe A, right camera view, carrying briefcase), and probe K (on grass, elapsed time). The silhouette in each frame over one walking cycle, of approximately 30–40 image frames was manually specified. This cycle was chosen to begin at the right heel strike phase of the walking cycle through to the next right heel strike. Whenever possible, this gait cycle was selected from the same 3D location in each sequence. In addition to marking a pixel as being from the background or subject, more detailed specifications in terms of body parts were marked. The head, torso, left arm, right arm, left upper leg, left lower leg, right upper leg, and right lower leg were explicitly labeled using different colors.

The Challenge Experiments

Along with the dataset, the gait challenge problem includes a definition of 12 challenge experiments (A–L),

![Evaluation of Gait Recognition. Figure 2 Partitioning of the HumanID gait challenge dataset in terms of its covariates, which are coded as follows: C – concrete surface, G – grass surface, A – first shoe type, B – second shoe type, BF – carryingabriefcase, NB – nobriefcase, M – data collected in May, N1 – new subjects in November data, and N2 – repeat subjects in November. The shaded cells are used to design the challenge experiments.](image)
spanning different levels of difficulty. This provides a common benchmark to compare performance with other algorithms. The experiments are designed to investigate the effect on performance of five factors, i.e., 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 gallery set is common for all the experiments and corresponds to the dark colored cell in Fig. 2. The gallery consists of sequences with the following covariates: Grass, Shoe Type A, Right Camera, No Briefcase, and collected in May along with those from the new subjects from November. This set was selected as the gallery because it was one of the largest for a given set of covariates. The experiments differ in terms of the probe sets, which are denoted by the lightly shaded cells. The structure of the 12 probe sets is listed in Table 1. The signatures are the video

Evaluation of Gait Recognition. Figure 3 Frames from (a) the left camera for concrete surface, (b) the right camera for concrete surface, (c) the left camera for grass surface, (d) the right camera for grass surface.

Evaluation of Gait Recognition. Figure 4 Top row shows the color images, cropped around the person, for one sequence. The bottom row shows the corresponding part-level, manually specified silhouettes.
sequences of gait. The last two experiments study the impact of elapsed time. The elapsed time covariate implicitly includes a change of shoe and clothing because the subjects were not required to wear the same clothes or shoes in both data collections. Because of the implicit change of shoe, it can be safely assumed that a different set of shoes were used in the May and November data collections. This is noted in Table 1 by A/B for shoe type in experiments K and L. The key experiments are those that involve controlled change in just one covariate and are marked with an asterisk in the table. The results from the 12 experiments provide an ordering of difficulty of the experiments.

**Baseline Gait Algorithm**

The third aspect of the gait challenge problem is a simple but effective baseline algorithm to provide performance benchmarks for the experiments. Ideally, this should be a combination of “standard” vision modules that accomplishes the task. Drawing from the success of template based recognition strategies in computer vision, a four-part algorithm that relies on silhouette template matching was designed. The first part semi-automatically defines bounding boxes around the moving person in each frame of a sequence. The second part extracts silhouettes from the bounding boxes using expectation maximization based on Mahalanobis distance between foreground and background color model at each pixel. Each silhouette is scaled to a height of 128 pixels and centered (automatically) in each frame along the horizontal direction so that the centerline of the torso is at the middle of the frame. The third part computes the gait period from the silhouettes. The gait period is used to partition the sequences for spatial-temporal correlation. The fourth part performs spatial-temporal correlation to compute the similarity between two gait sequences.

Let $S_P = \{S_P(1), \ldots, S_P(M)\}$ and $S_G = \{S_G(1), \ldots, S_G(N)\}$, be the probe and the gallery silhouette sequences, respectively. First, the probe (input) sequence is partitioned into subsequences, each roughly

**Evaluation of Gait Recognition.**

<table>
<thead>
<tr>
<th>Exp.</th>
<th>Probe</th>
<th>Number of subjects</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>(G, A, L, NB, M + N)</td>
<td>122</td>
<td>V^6</td>
</tr>
<tr>
<td>B</td>
<td>(G, B, R, NB, M + N)</td>
<td>54</td>
<td>S^7</td>
</tr>
<tr>
<td>C</td>
<td>(G, B, L, NB, M + N)</td>
<td>54</td>
<td>S + V</td>
</tr>
<tr>
<td>D</td>
<td>(C, A, R, NB, M + N)</td>
<td>121</td>
<td>F^8</td>
</tr>
<tr>
<td>E</td>
<td>(C, B, R, NB, M + N)</td>
<td>60</td>
<td>F + S</td>
</tr>
<tr>
<td>F</td>
<td>(C, A, L, NB, M + N)</td>
<td>121</td>
<td>F + V</td>
</tr>
<tr>
<td>G</td>
<td>(C, B, L, NB, M + N)</td>
<td>60</td>
<td>F + S + V</td>
</tr>
<tr>
<td>H</td>
<td>(G, A, R, BF, M + N)</td>
<td>120</td>
<td>B^9</td>
</tr>
<tr>
<td>I</td>
<td>(G, B, R, BF, M + N)</td>
<td>60</td>
<td>S + B</td>
</tr>
<tr>
<td>J</td>
<td>(G, A, L, BF, M + N)</td>
<td>120</td>
<td>V + B</td>
</tr>
<tr>
<td>K</td>
<td>(G, A/B, R, NB, N)</td>
<td>33</td>
<td>T^1 + S + C^9</td>
</tr>
<tr>
<td>L</td>
<td>(C, A/B, R, NB, N)</td>
<td>33</td>
<td>F + T + S + C</td>
</tr>
</tbody>
</table>

The gallery for all of the experiments is (G, A, R, NB, M + N) and consists of 122 individuals.

^6Key experiments  
^7View  
^8Shoe  
^9Surface  
^10Carry  
^11Elapsed time  
^12Clothing
over one gait period, \( N_{\text{Gait}} \). Gait periodicity is estimated based on periodic variation of the count of the number of foreground pixels in the lower part of the silhouette in each frame over time. 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).

Second, each of these probe subsequences, \( S_{pk} = \{ S_p(k), \ldots, S_p(k + N_{\text{Gait}}) \} \), is cross correlated with the given gallery sequence, \( S_G \).

\[
\text{Corr}(S_{pk}, S_G)(l) = \sum_{j=1}^{N_{\text{Gait}}} S(S_p(k+j), S_G(l+j)),
\]

where, the similarity between two image frames, \( S(S_p(i), S_G(j)) \), is defined to be the Tanimoto similarity between the silhouettes, i.e., the ratio of the number of common pixels to the number of pixels in their union. The overall similarity measure is chosen to be the median value of the maximum correlation of the gallery sequence with each of these probe subsequences. The strategy for breaking up the probe sequence into subsequences allows the algorithm to overcome segmentation errors in some contiguous sets of frames due to some background subtraction artifact or due to localized motion in the background.

\[
\text{Sim}(S_p, S_G) = \text{Median}_l \left( \max \text{Corr}(S_{pk}, S_G)(l) \right).
\]

The baseline algorithm is parameter free. The algorithm, although straightforward, performs quite well on some of the experiments and is quite competitive with the first generation of gait recognition algorithms.

**Performance on the Gait Challenge Problem**

The results reported for the Gait Challenge problem are of two types, ones that report results on the first version of the dataset that was released with 71 subjects and the second set of results are those reported for the full dataset with 122 subjects. The smaller dataset allows just the first eight experiments listed in Table 1, but with reduced gallery set sizes. Figure 5a tracks the baseline performance and the best performance reported in the literature. As of middle of 2007, there were 18 papers that reported results on the smaller version of the problem.

In 2002, when the Gait Challenge Problem was released, the performance of the baseline algorithm was better than the best reported performance. By 2004, while the baseline algorithm performance improved as the algorithm was fine-tuned, the performance of the best performance improved significantly and continued to improve through 2006. This trend is also seen in the results reported in six papers on the full dataset, summarized in Fig. 5b.

As is evident, the gait challenge problem has already spurred the development of gait recognition algorithms with improving performance. What is particularly interesting to notice is that the performance on hard experiments such those across surface (experiment D) and elapsed time (experiment K) has improved. Of course, there is still room for further improvement. Another interesting aspect is that the improvement of performance from 2004 to 2006 was not due to “continued engineering” of existing approaches, but involved the redesign of the recognition approaches. For instance, the greatest gains came from approaches that analyzed the silhouette shapes rather than the dynamics [2, 7]. Dynamics is important, but by itself is not sufficient.

Performance of a large number of algorithmic approaches have been explored. A review of the performances reported in these works reveals* [8] that

1. All most all of these approaches are based on silhouettes.
2. There is no one method that performs the best on all experiments.
3. Performances that involve matching against viewpoint and shoe variations, but on the same walking surface, has plateaued out.
4. Matching against walking surface variation remains a challenge.
5. Good performances (> 80%) has been reported for matching with and without carrying objects.
6. Matching across 6 months time difference has low performance, but the number of subjects involved in this experiment (33 subjects) is too low to derive meaningful conclusions.

**Other Large Datasets**

There are currently two other datasets that are as large as the HumanID gait challenge dataset in terms of number of subjects. First is the CASIA Indoor Gait Database [4].
The gallery set includes 124 subjects with normal walking, no coat and no carry bag. Different probes can be defined in terms of changes in (1) viewpoint, (2) clothing change (coat vs. no-coat), and (3) carrying a bag and not carrying a bag. Not many algorithms have reported performance on this dataset yet. But, the performance reported for the gait energy image approach in [9], seems to corroborate the findings from Figure 5.

**Evaluation of Gait Recognition.** Figure 5 Improvement in gait recognition algorithms over time with respect to the baseline performance. (a) Results on the first release of the gait dataset with 71 subjects in the gallery for the first eight experiments listed in Table 1 are tracked here. (b) Results on the full dataset with 122 subjects for the key experiments listed in Table 1. From 2004 to 2006, the best reported performances are better on all the experiments.
the HumanID Problem for matching across carrying
conditions: a performance of up to 80% is reported for
the CASIA dataset.

The other large dataset is the SOTON HID Gait
Database [10] with 115 subjects are collected mostly
indoor and some under outdoor conditions. The in-
door SOTON dataset was collected to examine the
premise that gait is unique so the background is con-
trolled so as to allow easy segmentation. The same
subjects were also filmed walking outdoors to deter-
mine whether gait biometrics could be perceived with
complex backgrounds. Performances in the range of
72–85% have been reported for matching across ses-
sions using a variety of approaches. This dataset also
affords the matching across time issue. It has been
shown that a time-dependent predictive model [11]
results in 92% recognition, but only for ten subjects.

It is worth noting that face recognition on
these data sets would be poor, indeed given the low-
resolution and the uncontrolled lighting.

Future Evaluations

It is to be expected that each gait research group
would collect their own data set to develop ideas.
This is an important process. For instance, one new
dataset is the CASIA infrared night gait dataset [12].
It consists of gait data from 153 subjects are collected
outdoors, at night, with and without carrying condi-
tion, and at two different speeds. This dataset nicely
complements existing datasets that are collected dur-
ing the day. Given the data-driven nature of biometrics
research, the key to future progress are such as data sets
collected to explore issues not considered or raised by
existing ones. For instance, as of today there is a need
for the better understanding of the variation of gait
due to surface conditions and across elapsed time.
Also, currently there is no dataset to explore the
matching across time issue for a large number of
subjects.

Ideally, the new datasets should consist of gait data
from around 1,000 subjects, an order of magnitude
larger than current large datasets. It is important to
increase the number of subjects so, that it is possible to
empirically study the scaling of performance with
number of subjects. Some guidance about the required
sizes can be found in [13, 14], where statistical
reasoning is employed to relate the number of subjects
with target error confidences. The data collection
should include gait data repeated at regular time inter-
vals of weeks, spanning about a year. The dataset
should be collected in outdoor conditions, preferably
collected at a distance of 300 m to reflect real world
conditions. The dataset should come with a set of well
defined experiments in terms of gallery and probe sets.
These experiments will influence the types of algo-
rithms. For the experiments to be effective at influen-
cing the direction of gait research the design of the
experiments needs to solve the three bears problem; the
experiments must be neither too hard nor too easy, but
just right. If performance on the experiments is easily
saturated, then the gait recognition community will
not be challenged. If experiments are too hard, then
it will not be possible to make progress on gait recog-
nition. Ideally, the set of experiments should vary in
difficulty, characterize where the gait recognition prob-
lem is solvable, and explore the factors that affect
performance. A set of experiments cannot meet this
ideal unless the appropriate set of data is collected. It is
important to view biometrics research as a data-driven
algorithm development process rather than algorithm-
driven data collection process.

Related Entries

- Hidden Markov Models
- Human Detection and Tracking
- Human Movement, Psychology
- Performance Evaluation, Overview
- Performance Testing Methodology, Standardization
- Psychology of Gait Recognition
- Silhouette-Based Gait Recognition
- Verification/Authentication/Identification/
Recognition

References

   Tech. rep CMU-RI-TR-01-18, Robotics Institute, Carnegie
   Mellon University (2001)
2. Kale, A., Sundaresan, A., Rajagopalan, A.N., Cuntoor, N.P.,
   of humans using gait. IEEE Transactions on Image Pro-
   and analysis of gait. In: International Conference on Automatic
   Face and Gesture Recognition, pp. 11–16 (2004)
**Author Query Form**

**Encyclopedia of Biometrics**  
**Alpha: E**

<table>
<thead>
<tr>
<th>Query Refs.</th>
<th>Details Required</th>
<th>Author's response</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU1</td>
<td>Please confirm if you are the copyright holder for the figure. If it is from another source, please confirm whether the necessary permissions have been obtained to reproduce the figure. Also, please provide us the permission details to be carried while citing the figure source.</td>
<td></td>
</tr>
</tbody>
</table>