Human identification at a distance via gait recognition

Liang Wang

Center for Research on Intelligent Perception and Computing (CRIPAC)
National Lab of Pattern Recognition (NLPR)
Institute of Automation, Chinese Academy of Sciences (CASIA)
# Outline

1. Introduction and overview

2. Traditional approaches for gait-based human identification
   - History and databases
   - Gait representation and learning algorithms

3. Deep networks for gait-based human identification
   - Cross-view gait based human identification with deep CNNs

4. How to build a practical gait-based human identification system?
   - End-to-end deep network for gait segmentation & recognition
   - System demo

5. Open questions and discussion
## Outline

1. Introduction and overview

2. Traditional approaches for gait-based human identification
   - History and databases
   - Gait representation and learning algorithms

3. Deep networks for gait-based human identification
   - Cross-view gait based human identification with deep CNNs

4. How to build a practical gait-based human identification system?
   - End-to-end deep network for gait segmentation & recognition
   - System demo

5. Open questions and discussion
What is Gait Recognition?

GAIT is a kind of behavioral biometric feature, whose raw data are video sequences presenting walking people. The goal of gait recognition is to identify people based on their gait features.

Movie “Mission Impossible 5”
Is gait recognition necessary?

- Short distance
  - Cooperative
    - Fingerprint
    - Iris
    - Face

- Long distance
  - Uncooperative
    - Gait
Is gait recognition necessary?

As a biometric, gait is still available at a distance when other biometrics are obscured or at too low resolution. Therefore, we need gait recognition.

Advantages: insensitive to distance, resolution, view, illumination
How does a gait recognition system work?

Intermediate representation
e.g. Gait Energy image

Database

Learning Algorithm

Human ID
Applications of gait recognition

- Suspect searching
- Access control
- Airport security
- Robotics & Smart home
## Outline

1. **Introduction and overview**
2. **Traditional approaches for gait-based human identification**
   - History and databases
   - Gait representation and learning algorithms
3. **Deep networks for gait-based human identification**
   - Cross-view gait based human identification with deep CNNs
4. **How to build a practical gait-based human identification system?**
   - End-to-end deep network for gait segmentation & recognition
   - System demo
5. **Open questions and discussion**
History of gait recognition:

- Aristotle (~350 BC): The first to analyze gait. “On the gait of animals”
- Leonardo da Vinci (~1500): movement sketches
- Borelli (1600s): Father of biomechanics, study the mechanical principles of locomotion. ‘De Motu Animalium’
History of gait recognition:

1600s

Shakespeare observed recognition:
• “High’st Queen of state; Great Juno comes; I know her by her gait” [The Tempest]
• “For that John Mortimer....in face, in gait in speech he doth resemble” [Henry IV/2]

Other literature: e.g. Band of Brothers: “I noticed this figure coming, and I realized it was John Eubanks from the way he walked”
History of gait recognition:

1800s

Eadweard Muybridge (1830-1904):

- Pioneering work in photographic studies of motion and motion-picture projection.
- Studied horses (1872): whether all four feet of a horse were off the ground at the same time while trotting
- Studied movement (1884)

Galloping horse, animated in 2006, using photos by Eadweard Muybridge

The Horse in Motion by Eadweard Muybridge. Running at a 1:40 pace. Frames 1-11 used for animation
History of gait recognition:

- Murray (1964): Produced standard movement patterns for pathologically normal people, suggesting the uniqueness of gait for individuals. ‘Walking Patterns of Normal Man’ ‘Gait As a Total Pattern of Movement’.

- Johansson (1973): Studied visual perception of motion patterns and suggested that ‘biological motion’ has far higher complexity than mechanical motions, and presented point-light displays to simulate human gait. ‘Visual Perception of Biological Motion and a Model for its Analysis’.

- Cutting & Kozlowski (1977): Announced that humans can recognize friends of a person solely by their gait with 70-80% accuracy. ‘Recognizing friends by their walk: Gait perception without familiarity cues’
History of gait recognition:

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>First gait biometrics paper - Cunado, Nixon and Carter (AVBPA 1997) - 90% CCR</td>
</tr>
<tr>
<td>2000</td>
<td>DARPA Program: Human ID at a distance</td>
</tr>
<tr>
<td>2015</td>
<td>Learning Representative Deep Features for Image Set Analysis, TMM</td>
</tr>
<tr>
<td>2016</td>
<td>Cross-view gait based human identification with deep CNNs, TPAMI</td>
</tr>
<tr>
<td></td>
<td>GEINET: view-invariant gait recognition, ICB</td>
</tr>
</tbody>
</table>

- **1997**: Design hand-crafted features for gait recognition
- **2015**: Deep learning for gait recognition
The DARPA program motivated the research on gait recognition.
# Released gait databases

Widely used benchmarks in the community

a) CASIA-B  

b) USF HumanID  

c) OU-ISIR, Large Population

<table>
<thead>
<tr>
<th>Name</th>
<th>Subjects</th>
<th>Sequences</th>
<th>Covariates</th>
<th>Viewpoints</th>
<th>Indoor(I)/Outdoor(O)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMU MoBo (30)</td>
<td>25</td>
<td>600</td>
<td>Y</td>
<td>6</td>
<td>I (Treadmill)</td>
</tr>
<tr>
<td>Georgia Tech (31)</td>
<td>15</td>
<td>268</td>
<td>Y</td>
<td>-</td>
<td>O</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>20</td>
<td>Y</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HID-UMD (32)</td>
<td>25</td>
<td>100</td>
<td>N</td>
<td>1</td>
<td>O</td>
</tr>
<tr>
<td>55</td>
<td></td>
<td>222</td>
<td>Y</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>SOTON Small Database (33)</td>
<td>12</td>
<td>-</td>
<td>Y</td>
<td>3</td>
<td>I</td>
</tr>
<tr>
<td>SOTON Large Database (34)</td>
<td>115</td>
<td>2,128</td>
<td>Y</td>
<td>2</td>
<td>I/O</td>
</tr>
<tr>
<td>SOTON Multimodal (35)</td>
<td>&gt;300</td>
<td>&gt;5,000</td>
<td>Y</td>
<td>12</td>
<td>I</td>
</tr>
<tr>
<td>SOTON Temporal (36)</td>
<td>25</td>
<td>2,280</td>
<td>Y</td>
<td>12</td>
<td>I</td>
</tr>
<tr>
<td>USF HumanID (23)</td>
<td>122</td>
<td>1,870</td>
<td>Y</td>
<td>2</td>
<td>O</td>
</tr>
<tr>
<td>CASIA A (37)</td>
<td>20</td>
<td>240</td>
<td>Y</td>
<td>3</td>
<td>I</td>
</tr>
<tr>
<td>CASIA B (38)</td>
<td>124</td>
<td>1,240</td>
<td>Y</td>
<td>11</td>
<td>I</td>
</tr>
<tr>
<td>CASIA C (39)</td>
<td>153</td>
<td>1,530</td>
<td>Y</td>
<td>1</td>
<td>O</td>
</tr>
<tr>
<td>OU-ISIR, Treadmill A (40)</td>
<td>34</td>
<td>612</td>
<td>Y</td>
<td>1</td>
<td>I (Treadmill)</td>
</tr>
<tr>
<td>OU-ISIR, Treadmill B (41)</td>
<td>68</td>
<td>2,764</td>
<td>Y</td>
<td>1</td>
<td>I (Treadmill)</td>
</tr>
<tr>
<td>OU-ISIR, Treadmill C (42)</td>
<td>200</td>
<td>200</td>
<td>Y</td>
<td>25</td>
<td>I (Treadmill)</td>
</tr>
<tr>
<td>OU-ISIR, Treadmill D (43)</td>
<td>185</td>
<td>370</td>
<td>N</td>
<td>1</td>
<td>I (Treadmill)</td>
</tr>
<tr>
<td>OU-ISIR, LP (44)</td>
<td>4,007</td>
<td>7,842</td>
<td>N</td>
<td>2</td>
<td>I</td>
</tr>
<tr>
<td>TUM-IITKGP (45)</td>
<td>35</td>
<td>850</td>
<td>Y</td>
<td>1</td>
<td>O</td>
</tr>
<tr>
<td>TUM-GAID (46)</td>
<td>305</td>
<td>3,370</td>
<td>Y</td>
<td>1</td>
<td>O</td>
</tr>
<tr>
<td>WOSG (47)</td>
<td>155</td>
<td>684</td>
<td>Y</td>
<td>8</td>
<td>O</td>
</tr>
</tbody>
</table>

[Makihara et al. 2015]
GEIs of two subjects under different conditions. The obtained GEIs are more noisy and of lower quality due to the complex backgrounds.
CASIA-B database

Details

<table>
<thead>
<tr>
<th>Indoor/Outdoor</th>
<th>indoor</th>
</tr>
</thead>
<tbody>
<tr>
<td># of subjects</td>
<td>124</td>
</tr>
<tr>
<td># of carrying/walking conditions</td>
<td>3</td>
</tr>
<tr>
<td># of viewpoints</td>
<td>11</td>
</tr>
</tbody>
</table>

Normal Walk  | Wearing Coats  | Carrying bags

Diagram of setup:
- Path with cameras lined up along it.
- Calibration mask at one end of the path.
- Network connections to computers labeled 1, 2, and 3.
- Direction of walking indicated.

Images of different viewpoints:
- 0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162°, 180°
OU-ISIR database, Large population dataset

<table>
<thead>
<tr>
<th>Details</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indoor/Outdoor</td>
<td>indoor</td>
</tr>
<tr>
<td># of subjects</td>
<td>4,007(v1), 4,016(v2)</td>
</tr>
<tr>
<td>Age range</td>
<td>1-94 years old</td>
</tr>
<tr>
<td># of walking conditions</td>
<td>1</td>
</tr>
<tr>
<td># of viewpoints</td>
<td>4 (55, 65, 75, 85)</td>
</tr>
<tr>
<td># of backgrounds</td>
<td>1</td>
</tr>
</tbody>
</table>

**Diagram:**
- Walking direction
- Observation angle
- Backgrounds: 55 deg, 65 deg, 75 deg, 85 deg
- Camera center

**Images:**
- Male
- Female
- Younger to Elder
Another super large database for gait recognition [C. Song, Y. Huang, et al.]

**CASIA-HT database (expected to be released early next year)**

<table>
<thead>
<tr>
<th>Details</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Indoor/Outdoor</td>
<td>outdoor</td>
</tr>
<tr>
<td># of subjects</td>
<td>1000</td>
</tr>
<tr>
<td># of carrying conditions</td>
<td>3</td>
</tr>
<tr>
<td># of walking conditions</td>
<td>2</td>
</tr>
<tr>
<td># of viewpoints</td>
<td>13 horizontal, 2 vertical</td>
</tr>
<tr>
<td># of backgrounds/scenarios</td>
<td>2</td>
</tr>
<tr>
<td># of sequences</td>
<td>&gt;760,000</td>
</tr>
</tbody>
</table>
## Categories of learning methods for gait recognition

<table>
<thead>
<tr>
<th>Model-based:</th>
<th>use the human body structure</th>
<th>Model free (appearance-based):</th>
<th>use the whole motion pattern of the human body</th>
</tr>
</thead>
</table>

![Diagram](image_url)

- **Images**
- **Profiles**
- **GEIs**
- **Learning**
  - PCA
  - LDA
  - LPP
- **Recognition**
  - SVM
  - NN
  - ...

**Human ID**

**Step-by-step**
## Categories of learning methods for gait recognition

<table>
<thead>
<tr>
<th>Model-based: use the human body structure</th>
<th>Model free (appearance-based): use the whole motion pattern of the human body</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Greater invariant properties and better at handling occlusion, noise, scale and rotation.</td>
<td>• Computational efficiency and simplicity</td>
</tr>
<tr>
<td>• Require a high resolution and are not yet very suitable for outdoor surveillance</td>
<td>• Can handle low-resolution case</td>
</tr>
<tr>
<td></td>
<td>• Suitable for outdoor surveillance</td>
</tr>
</tbody>
</table>
Model-based approaches: an example

• Fusion of static and dynamic body information.

• The static body information is in a form of a compact representation obtained by Procrustes shape analysis.

• The dynamic information is obtained by a model based approach which tracks the subject and recover joint-angle trajectories of lower limbs.

• Fusion at the decision level used to improve recognition results.

Fusion of Static and Dynamic Body Biometrics for Gait Recognition, Liang Wang, Huazhong Ning, Tieniu Tan, Weiming Hu, ICCV 2003
Model-free approaches: examples


• TSVD: Multiple views gait recognition using view transformation model based on optimized gait energy image,” in Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences (THEMIS), 2009.


(Intermediate) Gait Representation

GEI
(Gait Energy Image)

GEI
(Gait Entropy Image)

GFI
(Gait Flow Image)

CGI
(Chrono Gait Image)

Most widely used
One key concept: gait cycle

- Between where the same foot touches the ground for the first and second time.
- For the purpose of normalization of silhouettes and computing gait templates such as GEI.
Gait Energy Image (GEI)

- Spatially well-aligned, temporally averaged gait frames within one gait cycle
- Empirically 30 frames/whole sequence of frames enough to cover a complete gait cycle.
- \( F(i, j) \) indicates how likely there appears part of a human body in the position \((i, j)\)
- GEI is robust to the silhouette noise, but may have a high dimensionality

\[
F(i, j) = \frac{1}{T} \sum_{t=1}^{T} I(i, j, t)
\]

Gait Entropy Image (GEnI)

- Calculate Shannon entropy for each pixel in the silhouette images.
- The dynamic area of human body (legs and arms) are represented by higher intensity values in the GEnIs. In contrast, the static areas such as torso give rise to low intensity values.
- Silhouette pixel values in the dynamic areas are more uncertain and thus more informative leading to higher entropy values.

\[ H(x,y) = -\sum_{k=1}^{K} p_k(x,y) \log_2 p_k(x,y) \]

\[ G(x,y) = \frac{(H(x,y) - H_{\text{min}}) \times 255}{(H_{\text{max}} - H_{\text{min}})} \]

GFI contains the motion information of the human gait. GFIs are generated by determining the optical flow field from the binary silhouettes of each cycle.

\[
(uF_t, i(x, y), vF_t, i(x, y)) = OpticalFlow(SI_t, i(x, y), SI_{t+1}, i(x, y))
\]

\[
MagF_t, i(x, y) = \sqrt{(uF_t, i(x, y))^2 + (vF_t, i(x, y))^2}
\]

\[
BF_t, i(x, y) = \begin{cases} 
0 & \text{if } MagF_t, i(x, y) \geq 1 \\
1 & \text{otherwise}
\end{cases}
\]

\[
GFI_t(x, y) = \frac{\sum_{t=1}^{N-1} BF_t, i(x, y)}{N}
\]
• A great advantage of using GFI is that the number of GFIs is smaller than the number of silhouette images. In other words, GFI is more computationally efficient.

• However, if the silhouettes are extracted at a low quality, a GFI may be embedded with irrelevant information, which affects the recognition rate.

We encode temporal information in the silhouette images with additional colors to generate a chrono-gait image.

The goal of CGIs is to compress the silhouette images into a single image without losing too much temporal relationship between the images.

Performance of different gait representations

A recent empirical study by Iwama et al. shows that GEI, despite of its simplicity, is the most stable and effective kind of features for gait recognition on their proposed dataset with 4,007 subjects.

### Outline

1. Introduction and overview

2. Traditional approaches for gait-based human identification
   - History and databases
   - Gait representation and learning algorithms

3. Deep networks for gait-based human identification
   - Cross-view gait based human identification with deep CNNs

4. How to build a practical gait-based human identification system?
   - End-to-end deep network for gait segmentation & recognition
   - System demo

5. Open questions and discussion
The pipeline of a typical GEI-based gait recognition method.

1. Extract human silhouettes from video sequences

2. Align and average the silhouettes along the temporal dimension to get a GEI.

3. Given a probe GEIs and those in the gallery, evaluate the similarities between each pair of probe and gallery GEIs.

4. Assign the identity of the probe GEI, usually with the nearest neighbor classifier.

Different from previous methods, here the third step above is realized with deep convolutional neural networks (CNN).

Z. Wu, Y. Huang, L. Wang, X. Wang, T. Tan, A comprehensive study on cross-view gait based human identification with deep CNNs, IEEE TPAMI, 2016
Robustness of gait recognition system

One of the biggest challenges is to disentangle the identity-unrelated factors

- subject-related ones: walking speed, dressing and carrying conditions,
- device-related ones: different frame rates and filming resolutions,
- environment-related ones: illumination conditions and camera viewpoints.

Among these, the change of viewpoints would be one of the most tricky factors.
The performance of an approach ignoring cross-view variations would drop drastically when the viewpoint changes. Because the appearances of objects can be substantially altered, leading to intra-class variations larger than inter-class variations.
Feature learning for gait recognition

Traditional learning pipeline

Deep learning pipeline

It is difficult to manually design view-invariant feature representations for gait recognition
A similarity learning approach

Class label

Neural Networks

0 1

Neural Networks

(Similarity learning)

- Few labeled multi-view human walking videos, many labeled pairs
- Train deep networks to recognize the most discriminative changes of gait patterns which suggest the change of human identity
Three network architectures to be investigated.
Network architectures

1) Matching Local Features at the Bottom Layer (LB)

- Pairs of GEIs are compared within local regions
- Only linear projection is applied before computing the differences between pairs of GEIs, which is realized by the sixteen pair-filters in the bottom-most convolution stage.
- A pair-filter takes two inputs and can be seen as a weighted comparator.
- At each spatial location, it will first re-weight the local regions of its two inputs respectively, and then render the sum of these weighted entries to simulate the subtraction.
The “Subtraction” pair-filters

- Some of the learned pair-filters are subtracting gallery GEIs from probe GEIs.
- Project GEIs of different views into a common space where the GEIs become more comparable.
- There are two more convolution stages above the matching layer, whose nonlinearity is supposed to be beneficial to learning complex patterns from the differences between GEI pairs.
Network architectures

2) Matching Mid-Level Features at the Top Layer (MT)

- Two extra non-linear projections are applied
- The motivation is to apply deep non-linear normalization to GEIs instead of the shallow linear one in LB.
- LB directly computes the weighted differences at the bottom layer (with local features), and then learns to recognize the patterns in the obtained differences with the rest two convolution layers.
- In contrast, MT learns mid-level features first, and then computes the weighted differences.
- Model complexities of LB and MT are consistent
Network architectures

3) Matching Global Features at the Top Layer (GT)

• Pairs of GEIs are compared with each other by learned global features.
• Two more fully-connected layers compared with Network MT.
• The weighted differences are computed from global features at Layers F4 and F4'. Each of them is the descriptions of a whole GEI, with only 1,024 entries, which is much more compact than those of Networks LB and MT.
Model complexity

- The model complexity of Network GT is higher than the previous two due to the use of fully-connected layers, which can lead to over-fitting depending on the size of training data.

- However, the advantage of this network is its compactness, which can lead to computational efficiency.
  - First, we can store in advance the output of Layer F4’ for all gallery GEIs.
  - Second, feed a probe GEI to the network once and obtain the output of Layer F4.
  - Finally, compare the two 1,024-dimensional features using Layer F5 and the two-way classifier.
A substitute model: Compact Mid-Level & Top (CMT)

• We do NOT compare Global @ Top (GT) in detail considering its less satisfactory performance.

• In our experiments, Network GT suffers from severe over-fitting, probably due to the small training dataset. However, we sometimes do favor its computational efficiency.

• As a compromise, we modify Network MT to obtain more compact features, which is the very third network compared here, i.e., Compact Mid-Level & Top (CMT).

• It amounts to use a larger stride in the third convolution stage. For example, when we use a stride of five, the resulting feature map will be in size $3 \times 5 \times 256$, with only 3,840 entries.
**Inspiration**: Two-Stream Convolutional Networks for Action Recognition in Videos [NIPS 2015]

Two-stream architecture for video classification: capture the complementary information on **appearance** from still frames and **motion** between frames.
Network architectures

4) Two-stream network

• Composed of two LB networks.
• The left stream takes a pair of GEIs as the input, which is the counter part of the stream processing still images in Zisserman’s network.
• The right stream takes a pair of chrono-gait images (CGIs) as the input, which is the counter part of the stream processing optical flow features.
Network architectures

4) Two-stream 3D CNN network

• Train a network with 3D convolutions in its first & second layers.

• **Training:** each time we feed it with a pair of sequence slices, each of which contains nine adjacent frames sampled from a gait sequence.

• **Testing:** we feed it with all frames of a sequence (nine by nine to fit the network input), and average the output.
Experimental results

Comparison of our method with previous ones on CASIA-B by average accuracies. Models are trained with GEIs of the first 24 subjects.

<table>
<thead>
<tr>
<th>Gallery NM #1-4</th>
<th>0°-180°</th>
<th>36°-144°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe NM #5-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVR [30]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSVD [29]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMCC [12]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ViDP [23]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ours</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>0°</th>
<th>54°</th>
<th>90°</th>
<th>126°</th>
<th>54°</th>
<th>90°</th>
<th>126°</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>29</td>
<td>34</td>
<td>35</td>
<td>44</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>33</td>
<td>42</td>
<td>49</td>
<td>50</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>46.3</td>
<td>52.4</td>
<td>48.3</td>
<td>56.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>59.1</td>
<td>50.2</td>
<td>57.5</td>
<td>83.5</td>
<td>76.7</td>
<td>80.7</td>
<td></td>
</tr>
<tr>
<td>54.8</td>
<td>77.8</td>
<td>64.9</td>
<td>76.1</td>
<td>90.8</td>
<td>85.8</td>
<td>90.4</td>
</tr>
</tbody>
</table>
Impact of network architectures

1) \( LB \approx MT \gg GT \): There are no significant gaps between the performances of LB and MT, and they both outperform GT with a clear margin.

2) \( LB \) vs. \( MT \): The most notable difference between the two is that MT performs better for view angles around 0° or 180°.

3) \( MT \) vs. \( CMT \): There is a moderate drop in performance for CMT compared with MT.

4) \( MT \) vs. \( Siamese \): The Siamese network can approximately be seen as a special case of MT.

5) \( 0° \approx 180° > 90° > \cdots > 36° \approx 144° \)
Influence of network depth
Influence of network depth

MT > LB ~ large LB > large MT > small LB.
Influence of network depth
Influence of network depth
Influence of input resolutions
Influence of input resolutions
Influence of data augmentation
Influence of input features
Influence of temporal information
Summary

Lack of Datasets for uncooperative gait recognition:
- A subject may halt, or turn around, so his/her gait sequence is not consecutive.
- There may be multiple subjects at the same time, and moving objects in the background, so it is harder to extract silhouettes.
- The cameras may be above the subjects, so more viewpoints should be considered.

It would be very hard to train cross-view gait recognition models on so small a dataset due to severe over-fitting.

Besides, considering the above mentioned factors, to re-identify a person in unscripted surveillance videos only relying on gait recognition, there still seems a long way to go. Probably, such a dataset with enough number of training data can push us forward to this goal.
Less heuristic preprocessing:

• There are many methods which can be used to improve our preprocessing. For example, pedestrian detection methods can locate a subject from complex backgrounds, pixel-wise labeling methods can extract silhouettes from raw images, and pose estimation methods can provide auxiliary information or help refining the silhouettes.

• Without these comprehensive methods, it would be intractable to deal with the above discussed kind of datasets for uncooperative gait recognition.

• But in this work, preprocessing is not our main concern, so we keep it as our future work.
# Outline

1. Introduction and overview

2. Traditional approaches for gait-based human identification
   • History and databases
   • Gait representation and learning algorithms

3. Deep networks for gait-based human identification
   • Cross-view gait based human identification with deep CNNs

4. How to build a practical gait-based human identification system?
   • End-to-end deep network for gait segmentation & recognition
   • System demo

5. Open questions and discussion
Images → Profiles → GEIs → Learning → Recognition

Step-by-step:

- Images
- Profiles
- GEIs
- PCA, LDA, LPP
- SVM, NN, ...
- Human ID

Less steps with CNN:

- Images
- Profiles
- GEIs
- CNN
- GaitNet
- Human ID

End-to-end:

- Images
- Profiles
- GEIs
- GaitNet
- Human ID
A simplest system

Method
- Background subtraction
- GEI template matching

Requirement
- Indoor
- Simple background and texture

Code
https://github.com/developfeng/GaitRecognition
An end to end gait recognition system

Images → Profiles → GEIs → Learning → Recognition

PCA, LDA, LPP → SVM, NN, ...

Step-by-step

Less steps with CNN

GaitNet → Human ID

End-to-end
Experimental analysis

CASIA-B

✧ Simple background;
✧ Indoor;
✧ Fine profiles.
## Experiments-Results on CASIA-B

<table>
<thead>
<tr>
<th>Method</th>
<th>NM</th>
<th>CL</th>
<th>BG</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GEI [1]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>0.9593</td>
<td>0.9355</td>
<td>0.8862</td>
<td>0.9270</td>
</tr>
<tr>
<td>LDA</td>
<td>1.0000</td>
<td>0.9839</td>
<td>0.9837</td>
<td>0.9892</td>
</tr>
<tr>
<td>LPP</td>
<td>1.0000</td>
<td>0.9758</td>
<td>0.9350</td>
<td>0.9703</td>
</tr>
<tr>
<td><strong>GENI [2]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>0.9675</td>
<td>0.9597</td>
<td>0.8943</td>
<td>0.9405</td>
</tr>
<tr>
<td>LDA</td>
<td>1.0000</td>
<td>0.9839</td>
<td>0.9675</td>
<td>0.9838</td>
</tr>
<tr>
<td>LPP</td>
<td>1.0000</td>
<td>0.9839</td>
<td>0.9431</td>
<td>0.9757</td>
</tr>
<tr>
<td><strong>GFI [3]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>0.9675</td>
<td>0.9516</td>
<td>0.9024</td>
<td>0.9405</td>
</tr>
<tr>
<td>LDA</td>
<td>0.9837</td>
<td>0.9113</td>
<td>0.9024</td>
<td>0.9325</td>
</tr>
<tr>
<td>LPP</td>
<td>0.8618</td>
<td>0.8065</td>
<td>0.7642</td>
<td>0.8108</td>
</tr>
<tr>
<td><strong>CGI [4]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>0.9512</td>
<td>0.9435</td>
<td>0.8943</td>
<td>0.9297</td>
</tr>
<tr>
<td>LDA</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9675</td>
<td>0.9892</td>
</tr>
<tr>
<td>LPP</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9512</td>
<td>0.9837</td>
</tr>
<tr>
<td><strong>GEI-CNN [5]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No-Joint</td>
<td>0.9756</td>
<td>0.9194</td>
<td>0.9024</td>
<td>0.9325</td>
</tr>
<tr>
<td>Joint</td>
<td>1.0000</td>
<td>0.9919</td>
<td>0.9839</td>
<td>0.9919</td>
</tr>
</tbody>
</table>
✧ Complex background;
✧ Outdoor;
✧ Hard to get profiles.
Visualization of Segmentation Network

Experiments-Joint Learning

SCENE-1

SCENE-2

NM

CL

BG

Image  Ground-Truth  Non-Joint  Joint  Image  Ground-Truth  Non-Joint  Joint
<table>
<thead>
<tr>
<th>Methods</th>
<th>S-1</th>
<th>S-2</th>
<th>S-3</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NM</td>
<td>CL</td>
<td>BG</td>
<td>NM</td>
</tr>
<tr>
<td>GEI [1]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>0.7971</td>
<td>0.8456</td>
<td>0.8623</td>
<td>0.9783</td>
</tr>
<tr>
<td>LDA</td>
<td>0.8841</td>
<td>0.8750</td>
<td>0.8623</td>
<td>0.9710</td>
</tr>
<tr>
<td>LPP</td>
<td>0.8696</td>
<td>0.8750</td>
<td>0.8913</td>
<td>0.9348</td>
</tr>
<tr>
<td>GEnl [2]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>0.7971</td>
<td>0.7868</td>
<td>0.7826</td>
<td>0.9855</td>
</tr>
<tr>
<td>LDA</td>
<td>0.8261</td>
<td>0.8603</td>
<td>0.8478</td>
<td>0.9710</td>
</tr>
<tr>
<td>LPP</td>
<td>0.8623</td>
<td>0.8603</td>
<td>0.8551</td>
<td>0.9348</td>
</tr>
<tr>
<td>GFI [3]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>0.8116</td>
<td>0.8382</td>
<td>0.8768</td>
<td>0.9565</td>
</tr>
<tr>
<td>LDA</td>
<td>0.7971</td>
<td>0.6838</td>
<td>0.8188</td>
<td>0.8841</td>
</tr>
<tr>
<td>LPP</td>
<td>0.6667</td>
<td>0.6985</td>
<td>0.7826</td>
<td>0.8188</td>
</tr>
<tr>
<td>CGI [4]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>0.7101</td>
<td>0.7299</td>
<td>0.8044</td>
<td>0.8696</td>
</tr>
<tr>
<td>LDA</td>
<td>0.7101</td>
<td>0.6861</td>
<td>0.7899</td>
<td>0.8478</td>
</tr>
<tr>
<td>LPP</td>
<td>0.7101</td>
<td>0.6861</td>
<td>0.7464</td>
<td>0.8406</td>
</tr>
<tr>
<td>GEI-CNN [5]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No-Joint</td>
<td>0.8623</td>
<td>0.9055</td>
<td>0.9348</td>
<td>0.9601</td>
</tr>
<tr>
<td>Joint</td>
<td>1.0000</td>
<td>0.9779</td>
<td>0.9816</td>
<td>0.9963</td>
</tr>
<tr>
<td>GaitNet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Demo of near-commercial gait recognition
<table>
<thead>
<tr>
<th>Outline</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction and overview</td>
</tr>
<tr>
<td>2. Traditional approaches for gait-based human identification</td>
</tr>
<tr>
<td>• History and databases</td>
</tr>
<tr>
<td>• Gait representation and learning algorithms</td>
</tr>
<tr>
<td>3. Deep networks for gait-based human identification</td>
</tr>
<tr>
<td>• Cross-view gait based human identification with deep CNNs</td>
</tr>
<tr>
<td>4. How to build a practical gait-based human identification system?</td>
</tr>
<tr>
<td>• End-to-end deep network for gait segmentation &amp; recognition</td>
</tr>
<tr>
<td>• System demo</td>
</tr>
<tr>
<td>5. Open questions and discussion</td>
</tr>
</tbody>
</table>
Future directions and open questions

• Multiple overlapping persons

• Soft biometrics: attributes based gait recognition: fast query retrieval

• Speedup of deep networks/ model learning

• Super large-scale gait databases: >10,000 subjects, real world scenarios

• Multi-modal human identification: face recognition + gait recognition
References


References

- W. Kusakunniran, Q. Wu, H. Li, and J. Zhang, “Multiple views gait recognition using view transformation model based on optimized gait energy image,” in Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences (THEMIS), 2009.
Acknowledgements

• Special thanks to Chunfeng Song and Zifeng Wu
• Members of Multi-Modal Computing Group & CRIPAC
THANK YOU!