Algorithms for Human Identification using a Monocular Video Sequence

Amit Kale
Center for Automation Research and Department of Electrical and Computer Engineering,
University of Maryland at College Park
Overview

- Introduction
- Gait-based human identification using appearance matching
- Statistical framework for gait-based human identification
- View Invariant Gait recognition
- Combining Multiple Evidences For Gait Recognition
- Future Research Directions
Introduction

- Automated Person Identification
- Use of Biometrics
  - Fingerprints, Hand Geometry, Face, Iris etc.
- Shortcomings of Traditional Biometrics
- Gait as a Biometric
- Uniqueness of Gait
Preprocessing

- Independence from Clothing, Illumination

Binarized silhouette
Appearance matching approach

- **Motivation**
  - Similarity with text-based speaker identification
  - Availability of limited training data

- **Feature**: Width of outer contour of silhouette
Width Feature

Width Vectors Overlay
Temporal plots of width

Person 1

Person 2

Person 3

Person 4
Spatio-temporal smoothing of width

\( \{W(1), \ldots, W(N)\}, W(i) \in \mathbb{R}^M \)

Eigen decomposition

\( \{V(1), \ldots, V(M)\} \)

\[ W'_r(i) = \left( \sum_{j=1}^{m} w_j V(j) \right) + \bar{W} \]

\[ w_j(i) = \langle W(i), V(j) \rangle, \bar{W} = \frac{W(1) + \ldots + W(N)}{N} \]
Width vectors overlays after smoothing

Before smoothing

After smoothing

Other features
- Direct Smoothed Width Vectors
- Dynamics
Matching Gait Sequences

- Template Matching Using DTW
  - Dynamic programming
  - Non-linear time normalization for matching
- Constraints
  - Monotonicity: $X_{k-1} \leq X_k$, $Y_{k-1} \leq Y_k$
  - Local continuity: $X_k - X_{k-1} = 1$, $Y_k - Y_{k-1} \leq 2$
  - Global path
  - End point: $X_T = Y_R$
DTW Algorithm

1. Local distance computation \( L(k,l) = ||Y_k - X_l|| \)

2. Cumulative distance computation
   \[
   D(X_k, Y_k) = L(X_k, Y_k) + \min \{ D(X_{k-1}, Y_k), D(X_{k-1}, Y_{k-1}), D(X_{k-1}, Y_{k-2}) \}
   \]

3. Backtracking
Experimental results

- Effect of Eigen Smoothing

<table>
<thead>
<tr>
<th>Feature\Rank</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvector 1</td>
<td>73</td>
<td>75</td>
<td>80</td>
<td>80</td>
<td>84</td>
</tr>
<tr>
<td>Eigenvectors 1,2</td>
<td>80</td>
<td>87</td>
<td>90</td>
<td>90</td>
<td>91</td>
</tr>
<tr>
<td>Eigenvectors 1,2,3</td>
<td>68</td>
<td>80</td>
<td>84</td>
<td>84</td>
<td>84</td>
</tr>
<tr>
<td>Eigenvectors 1,2,3,4</td>
<td>73</td>
<td>77</td>
<td>84</td>
<td>84</td>
<td>84</td>
</tr>
<tr>
<td>Eigenvectors 1,2,3,4,5</td>
<td>70</td>
<td>73</td>
<td>79</td>
<td>82</td>
<td>84</td>
</tr>
</tbody>
</table>

- Effect of speed

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Feature</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slow vs Slow</td>
<td>Direct-smoothed feature</td>
<td>70.8</td>
</tr>
<tr>
<td></td>
<td>Eigenv-smoothed feature</td>
<td>95.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95.8</td>
</tr>
<tr>
<td>Fast vs Fast</td>
<td>Direct-smoothed feature</td>
<td>83.3</td>
</tr>
<tr>
<td></td>
<td>Eigenv-smoothed feature</td>
<td>95.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100.0</td>
</tr>
<tr>
<td>Fast vs Slow</td>
<td>Direct-smoothed feature</td>
<td>54.1</td>
</tr>
<tr>
<td></td>
<td>Eigenv-smoothed feature</td>
<td>75.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>83.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>83.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>87.5</td>
</tr>
</tbody>
</table>
Warping Path
## USF Dataset

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Probe</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>G,A,L</td>
<td>View</td>
</tr>
<tr>
<td>B</td>
<td>G,B,R</td>
<td>Shoe</td>
</tr>
<tr>
<td>C</td>
<td>G,B,L</td>
<td>Shoe, View</td>
</tr>
<tr>
<td>D</td>
<td>C,A,R</td>
<td>Surface</td>
</tr>
<tr>
<td>E</td>
<td>C,B,R</td>
<td>Surface, Shoe</td>
</tr>
<tr>
<td>F</td>
<td>C,A,L</td>
<td>Surface, View</td>
</tr>
<tr>
<td>G</td>
<td>C,B,L</td>
<td>Surface, Shoe, View</td>
</tr>
<tr>
<td>H</td>
<td>G,A,R,BF</td>
<td>Briefcase</td>
</tr>
<tr>
<td>I</td>
<td>G,B,R,BF</td>
<td>Shoe, Briefcase</td>
</tr>
<tr>
<td>J</td>
<td>G,A,L,BF</td>
<td>View, Briefcase</td>
</tr>
<tr>
<td>K</td>
<td>G,A,R,t2</td>
<td>Time</td>
</tr>
<tr>
<td>L</td>
<td>C,A,R,t2</td>
<td>Surface, Time</td>
</tr>
</tbody>
</table>
Results on the USF database

![Bar chart showing identification rate (%) for different probes (A-G) using different methods: Baseline, DTW using Width Feature, and DTW using Binary Silhouette.](chart.png)

- A: Baseline 85%, DTW using Width Feature 80%, DTW using Binary Silhouette 75%
- B: Baseline 75%, DTW using Width Feature 70%, DTW using Binary Silhouette 65%
- C: Baseline 65%, DTW using Width Feature 60%, DTW using Binary Silhouette 55%
- D: Baseline 55%, DTW using Width Feature 50%, DTW using Binary Silhouette 45%
- E: Baseline 45%, DTW using Width Feature 40%, DTW using Binary Silhouette 35%
- F: Baseline 35%, DTW using Width Feature 30%, DTW using Binary Silhouette 25%
- G: Baseline 25%, DTW using Width Feature 20%, DTW using Binary Silhouette 15%
Statistical Framework for gait

Components of gait: Structure and dynamics

- **Features**
  - Width of the outer contour of the silhouette (UMD, CMU, USF)
  - Entire binary silhouette (USF)
Exemplars: Structure

- Distinct Stances occur during a walk cycle

- Divide gait cycles into \( N \) segments

- Pool features from the jth segment

- Pick \( e_j \) such that \( D_j = \sum_{i=1}^{M} \min_{j \in \{1, ..., N\}} d(x_i, e_j) \) is minimized

- Optimum Exemplar set \( \{e_1, ..., e_N\} \)

- Choice of \( N \)
Dynamics

- Difficulties with the simple classification criterion

\[
\{U_1, \ldots, U_P\} \quad \{E^1\} \quad \{E^j\} \quad \{E^N\}
\]

- Use dynamics of transition across exemplars

\[
A = \left[ p(e_i(t) \mid p(e_j(t)) \right]
\]
Problem: Dimensionality vs Training data available

Solution:
Indirect Approach:

- FED vectors \( f_j(t) = d(x^j(t), e_l) \) where \( t=1,\ldots,T, l=1,\ldots,5 \)
  - Encodes structure (\( D_u^i < D_j^u \)) and dynamics of individual
  - FED vector sequence as the observed process corresponding to the Markov matrix: HMM \( \lambda = (A, B, \Pi) \)
  - Generality of FED vector for different Representations
Training

Frame to Exemplar Distance (FED) Vector Sequences

Hidden Markov Model
Evaluation

<table>
<thead>
<tr>
<th>HMM #1</th>
<th>HMM #2</th>
<th>HMM #N</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_{i1}(t) )</td>
<td>( f_{i2}(t) )</td>
<td>( f_{iN}(t) )</td>
</tr>
</tbody>
</table>

Database of Exemplars

Video of unknown person "U"

\( d(x^n(t), e_i^n) \)

Rank order Pi’s

\( P_1 \) \( P_2 \) \( P_N \)
- **Direct Approach**
  - Usual Approach for HMMs: Mixture of Gaussians for modeling B.
  - Redefine B in terms of Exemplars

\[
b_n(x(t)) = P(x(t) \mid e_n) = \beta e^{-\alpha D(x(t), e_n)}
\]
Training

- We start with a predefined value for $A$, a uniform distribution for $\pi$, and the initial estimate of the exemplars.
- The Expectation-Maximization algorithm is used to refine the estimates of the exemplars and $A$.
- The model parameters usually converge in a few iterations.

**Updating Exemplars**

$$E_j^{(i+1)} = \arg_E \max \prod_{t \in \{j^{th}\ \text{group}\}} P(O_t \mid E) \Rightarrow E_j^{(i+1)} = \arg_E \min \sum_{t \in \{j^{th}\ \text{group}\}} D(O_t, E)$$

**Updating Transition Matrix, $A$**

$$A^{(i+1)} = \arg_A \max P(O \mid (A^{(i)}, B^{(i)}, \pi)) \quad (\text{Baum} – \text{Welch Algorithm})$$
Testing

- A sequence $X$ can be identified by finding the HMM parameters $(\lambda_p)$ from the gallery that maximizes the probability of the observation sequence given $\lambda_p$.
- We use the Viterbi algorithm to compute the probability of a sequence given the model.

$$ID = \arg_p \max P(X | \lambda_p),$$

where $\lambda_p$ is the HMM for $p^{th}$ person.
Results on the USF Database
Results on the USF Database

![Graph showing cumulative match scores for different probes across ranks.]

![Bar chart comparing baseline and HMM (image based) identification rates for different subjects.]
View Invariant Gait Recognition

- Limitations of present gait recognition algorithms
  - Require exact side-view of the walking person
  - Solution: 3D models (Hard!)
  - Alternative: Visual hull
    - Needs at least 4 cameras
    - Computation of the order of $O(kmn)$

- Idea: Person walking far from the camera can be approximated as a planar object
Overview of our method

Done entirely in video domain, no explicit 3D computation
Imaging Geometry

Translational velocity \([v_X, 0, v_Z]\)

\[Z_1 \gg f\]
Framework for novel view synthesis

- **Tracking**
  - Assume initial position of a fixed point on the object \((x_{ref}, y_{ref})\)

Tracks of \((x, y)\) positions of the head for different \(\theta\)

Slope of the lines = \(\tan(\alpha)\)
Robust Estimation of $\alpha$

LS estimate

LMEDS estimate
Estimation of $\theta$ (for constant velocity models)

$$\cot(\alpha) = \frac{x_{ref} - f \cot(\theta)}{y_{ref}}$$

Synthesis

$$x_0 = f \left( \frac{x_\theta \cos(\theta) + x_{ref} (1 - \cos(\theta))}{-\sin(\theta)(x_\theta + x_{ref}) + f} \right) + f$$

$$y_0 = f \left( \frac{y_\theta}{-\sin(\theta)(x_\theta + x_{ref}) + f} \right) + f$$
Synthesis Examples

15 degrees

30 degrees

45 degrees
Gait Recognition Results

- Feature: Binarized silhouette
- Classifier: DTW with binary correlation as local distance

\[
\theta = 15 \quad \theta = 30 \quad \theta = 45
\]
NIST database
Gait Recognition (NIST database)
Application: Multimodal Biometrics

Video of unknown person far from camera

Gait Recognition Algorithm

Obtain top “M” matches

Face Recognition Algorithm

Fusion

Identification
Fusion of Face and Gait

FACE

NIST Database

GAIT

Center for Automation Research, University of Maryland, College Park
Histogram of True Matches and False Matches
Fusion of Face and Gait

A. Hierarchical Fusion: Gait -> Face
   - top matches above a threshold
   - 1/5 th time for face recognition.

B. Product rule on similarity scores
   - 100 % recognition.
Application: Video-based rendering

- Video compression, indexing and retrieval
- Comparison with IBR approaches
- Compensation of non-planarity
  For $Z \gg a, b, X$ \[ W_a = \frac{f}{Z} (a \cos(\theta) + b \sin(\theta)) \]
- Given “a” and $\theta$, render only the portion corresponding to \[ W_t = \frac{f}{Z} a \cos(\theta) \]
\[ d(P', Q') = a\cos(\theta) + b\sin(\theta) \]

\[ \theta \]

Image Plane

\[ f \]

\[ Z \approx f, a, b, x \]
Images at 0

Before

After

15

30

45

Determination of non-planar regions
Fusion of Multiple Evidences

- Left and Right Projection Vectors

- Foot Dominance

<table>
<thead>
<tr>
<th>Feature</th>
<th>Rank</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set of 4 half cycles 1(a)</td>
<td></td>
<td>68.1</td>
<td>77.2</td>
<td>84.0</td>
<td>84.0</td>
<td>84.0</td>
</tr>
<tr>
<td>Set of 4 half cycles shifted by one 1(b)</td>
<td></td>
<td>70.4</td>
<td>79.5</td>
<td>81.8</td>
<td>86.3</td>
<td>86.3</td>
</tr>
<tr>
<td>Minimum of 1(a) and 1(b)</td>
<td></td>
<td>79.0</td>
<td>81.4</td>
<td>83.2</td>
<td>86.0</td>
<td>86.0</td>
</tr>
</tbody>
</table>

- Fusion of Frontal + Side Gait
Applications and Future Work

- Short-time Verification problems.
- Using “generalized” gait eigen vectors for subspace based activity recognition.
- Extensions of the view invariant approach using 2 cameras.
- 3D parameterized models for gait vs. 2D approaches.
- Applications in video indexing and retrieval.
- Using 3-D models of objects for synthesis of non-planar object.
  - Novel view synthesis and recognition of face images.
Acknowledgements

Amit RoyChowdhury, Prof. A.N. Rajagopalan, Prof. B. Yegnanarayana, Naresh Cuntoor, Aravind Sundaresan, Ashok Veeraraghavan, Narayanan Ramanathan