Tutorial Outline

• Introduction
• PART I: Face-based Human Recognition
• PART II: Gait-based Human Recognition
  – 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
• Discussions
Silhouette Based Approaches – Prior Work

**Model Based Methods**

- Structural
  - Stride parameters
  - Human parameters
  - Joint trajectories
- Joint Angles
- Dual Oscillator
- Articulated Model
- Linked Feature Trajectories
- Modeled

**Model Free Analysis**

- Area, point distribution Models, symmetry, relational statistics
- Unwrapped silhouette, silhouette similarity, key frame analysis
- Eigenspace Sequences
- Hidden Markov Model
- Gait Style and Content
- Average Silhouette, moments, higher order correlation, video oscillations
- Kinematic Features, Ellipsoidal Fits

**Moving Shape**

**Shape + Motion**
Gait Data Bases

- HumanID data base (USF/NIST) (1870 sequences from 122 subjects)
  - For each subject, two views, two surface types and two types of shoes. Some carried brief cases; some were imaged after 6 months.
- UMD (Two data sets: 25 subjects and 55 subjects)
- University of Southampton (Soton) (116 subjects)
## 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>
UMD Infrastructure

- 4 Cameras, 4.5m
- 1 Camera, 6m
- 1 Camera, roof
- 1 Video server
- N clients

Diagram:
- Multi-Cast Video Server
- Clients
- Camera 4, 4.5 m height, used
- Camera 5, 6 m height, used
- Camera 6, roof top camera
UMD Dataset Acquisition
UMD Integrated Software System

- Video Input
- Image Processing
- Background Subtraction
- Blob Tracking
- Contour Computation
- Gait Recognition
- Face Recognition
- Fusion
Preprocessing

- Independence from Clothing, Illumination

background subtraction

Binarized silhouette
UMD Background Subtraction Results

(a)

Stances along segment 1

(b)

Stances along segment 2

Stances along segment 3

Stances along segment 4
Tutorial Outline

• Introduction
• PART I: Face-based Human Recognition
• PART II: Gait-based Human Recognition
  – 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
• Discussions
Appearance Matching Approach

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

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

Person 1  Person 2  Person 3  Person 4

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

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

\[ W_r(i) = (\sum_{j=1}^{m} w_j V(j)) + \overline{W} \]

\[ w_j(i) = \langle W(i), V(j) \rangle, \overline{W} = \frac{W(1) + \ldots + W(N)}{N} \]
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 \( X_{k-1} \leq X_k, \quad Y_{k-1} \leq Y_k \)
    • Local continuity \( X_k - X_{k-1} = 1, \quad Y_k - Y_{k-1} \leq 2 \)
    • Global path
    • End point \( X_T = Y_R \)
Dynamic Time Warping (DTW) Algorithm

- Local distance computation $L(k,l) = | |Y_k - X_l| |
- Cumulative distance computation
  
  $D(X_k, Y_k) = L(X_k, Y_k) + \min\{ D(X_{k-1}, Y_k), D(X_{k-1}, Y_{k-1}), D(X_{k-1}, Y_{k-2}) \}$

- Backtracking
Results on the USF database

![Graph showing identification rate for different probes A to G, comparing Baseline, DTW using Width Feature, and DTW using Binary Silhouette.]
Tutorial Outline

• Introduction
• PART I: Face-based Human Recognition
• PART II: Gait-based Human Recognition
  – 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
• Discussions
Statistical Framework for Gait-based Human Identification

- 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 $j^{th}$ segment

$$D_j = \sum_{i=1}^{M} \min_{j \in \{1,...,N\}} d(x_i, e_j)$$

- Optimum Exemplar set $\{e_1, ..., e_N\}$
- Choice of $N$
Dynamics

• Difficulties with the simple classification criterion

\[ \{ U_1, \ldots, U_p \} \]

Quantize

\[ \{ E^1 \} \quad \{ E^j \} \quad \{ E^N \} \]

\[ D_1 \quad D_j \quad D_N \]

• Use dynamics of transition across exemplars

\[ A = [p(e_i(t) \mid p(e_j(t)))] \]
Hidden Markov Model (HMM)

- 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 (\( D_u^u < D_j^u \)) 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
Evaluations

Database of Exemplars

Video of unknown person “U”

$P_N$ $P_2$ $P_1$

HMM #N

HMM #2

HMM #1

Rank order Pi’s
Direct Approach

– Usual Approach for HMMs: Mixture of Gaussians for modeling B.
– Redefine B in terms of Exemplars

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

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

Updating Transition Matrix, $A$

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

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

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

- Gait Energy Image (GEI) is used as a feature to tackle silhouette errors.
- Use real Silhouettes with a distortion model to generate synthetic templates; Synthetic Templates account for gait in varying conditions.
- PCA and MDA features are fused to obtain recognition results.

Figure 1. System diagram of human recognition using proposed statistical feature fusion approach.
Gait Dynamics Normalization

- The dynamics of the observed probe sequences are normalized using the pHMM model.
- Population Hidden Markov Model is used as a Generic Model for walking.
- Viterbi Decoding for Recognition.
Some New Results from USF

From IEEE TPAMI, June 2006, Liu and Sarkar
Summary of the Top Rank Recognition for Experiments

From IEEE TPAMI, June 2006, Liu and Sarkar
Tutorial Outline

• Introduction
• PART I: Face-based Human Recognition
• PART II: Gait-based Human Recognition
  – 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
• Discussions
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 \([v_X, 0, v_Z]\)
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 = \(\tan (\alpha)\)
Framework for Novel View Synthesis

- Robust Estimation of $\alpha$

![LS estimate](image1)

![LMEDS estimate](image2)
Framework for Novel View Synthesis

- Estimation of $\theta$ (for constant velocity models)

\[
\cot(\alpha) = \frac{x_{\text{ref}} - f \cot(\theta)}{y_{\text{ref}}}
\]

- Synthesis

\[
x_0 = f \frac{x_\theta \cos(\theta) + x_{\text{ref}} (1 - \cos(\theta))}{-\sin(\theta)(x_\theta + x_{\text{ref}}) + f}
\]

\[
y_0 = f \frac{y_\theta}{-\sin(\theta)(x_\theta + x_{\text{ref}}) + f}
\]
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 & Walking Pattern
Gait Recognition (NIST Database)
Tutorial Outline

- Introduction
- PART I: Face-based Human Recognition
- PART II: Gait-based Human Recognition
  - 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
- Discussions
Application: Multimodal Biometrics

- Video of unknown person far from camera
- Portion of video of unknown person close to camera
- Gait Recognition Algorithm
  - Obtain top “M” matches
- Face Recognition Algorithm
  - Identification
- Fusion
Fusion of Face and Gait

FACE

NIST Database

GAIT

Similarity Matrix for Face

Cumulative Match Score

Rank

0

5

10

15

20

25

30

0

10

20

30

0

5

10

15

20

25

30

0

5

10

15

20

25

30

0.00

0.01

0.02

0.03

0.04

0.05

0.06

Similarity Matrix for Gait

Cumulative Match Score

Performance using synthesized images

Performance using unnormalized images

Rank

Rank
Fusion of Face and Gait

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.
Shape or Dynamics (or Is It Both?)

- Human perception
- Most gait recognition algorithms are shape based!
- Relative importance of shape and dynamics
- Definition of shape
  - “Shape is all the geometric information that remains when location, scale and rotational effects are filtered out from the object”.
  - Kendall’s Statistical Shape Theory used for the characterization of shape.
  - Pre-shape accounts for location and scale invariance alone.
Pre-Shape

- k landmark points (complex vector)
- Translational invariance: Subtract mean
- Scale invariance: Normalize the scale

\[ Z_c = \frac{CX}{\| CX \|} , \quad \text{where} \quad C = I_k - \frac{1}{k} 1_k 1_k^T \]
Feature Extraction

- Silhouette
- Landmarks
- Centered Landmarks
- Pre-shape vector
Distance Between Shapes

- Shape lies on a spherical manifold.

- Shape distance must incorporate the non-Euclidean nature of the shape space.

  1) Full Procrustes distance.
  2) Partial Procrustes distance.
  3) Procrustes distance.
Full Procrustes Distance

- Procrustes Fit

\[ d(Y, X) = \left\| \beta - \alpha s e^{i\theta} - (a + jb)1_k \right\|. \]

- Full Procrustes distance = Minimum Procrustes fit.

\[ d_F(Y, X) = \inf_{s, \theta, a, b} d(Y, X). \]
Other Shape Distances

- Partial procrustes distance
  \[ d_P(X,Y) = \inf_{\Gamma \in SO(m)} \| \beta - \alpha \Gamma \|. \]

- Procrustes distance \((\rho)\): distance on the Great circle.
  \[ d_F(X,Y) = \sin \rho, \]
  \[ d_P(X,Y) = 2 \sin(\frac{\rho}{2}). \]
Tangent Space

• Linearization of spherical shape space around a particular pole.

• The Procrustes mean shape is usually chosen as the pole.

• If the shapes in the data are very close to each other then Euclidean distance in tangent space approximates shape distances.
Three Shape Based Methods for Recognition

• Stance Correlation.

• Dynamic time warping in shape space.

• Hidden Markov Model in shape space.
Stance Correlation

• Exemplars for 6 stances for each individual.

• The correlation between exemplars is used as the matching criterion.

• Performance comparable to Baseline.
**Dynamic Time Warping in Shape Space.**

- Enforce end-point constraint.
- Obtain best warping path.
- Cumulative error is computed using the shape distances described.
- Performance is better than baseline.
Hidden Markov Model in Shape Space

- Exemplars are regarded as states.
- HMM built for each person in the gallery.
- Identity established by maximizing the probability that the observation came from the model in the gallery.
- Performance is better than baseline and comparable to DTW.
Comparison of Various Methods on the USF Database

![Comparison of average (of Probes A-G) CMS curve for various algorithms](image-url)
Comparison of Various Methods on the USF Database

- Shape is more important for recognition than dynamics. Shape also provides for speed change invariance.
- Dynamics can help to improve performance of shape based methods.
Tutorial Outline

• Introduction
• **PART I: Face-based Human Recognition**
• **PART II: Gait-based Human Recognition**
  – 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**
• Discussions
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.
Applications and Future Work

• Is gait effective?
  – Maybe for a small data set (< 100 persons) viewed from fronto-parallel direction.
  – Can be fooled by changing the shoe type, intentional disguises etc.
  – Starbucks 8:00 a.m. gait versus going home gait!

• Gait analysis is useful for detecting abnormal walking patterns
References

- Numerous journal and conference papers from UMD, U. of Southampton, MIT, CMU, USF....
Tutorial Outline

• Introduction
• PART I: Face-based Human Recognition
• PART II: Gait-based Human Recognition
• Discussions