Semantic (Biometric) Recognition, Learning and Retrieval

Mark S. Nixon
University of Southampton UK
Acknowledgements

• Sina Samangooei, University of Southampton
• T. Raju Damarla, U.S. Army Research Laboratory
• Baofeng Guo, Hangzhou University
• Funding via Information Technology Alliance (MoD/ ARL)
Overview

• Can recognise **people** by measurements of **gait**
• Can recognise **vehicles** by measurements of **sound**
• Can recognise each by **semantic descriptions**
• **Fuse** semantic descriptions to **improve** recognition
• **Recall** objects by semantic description
• Can **learn semantic labels** from data

**Contribution:** use semantics to enrich fusion, classification and retrieval
Surveillance

- Increased interest in surveillance technologies
  - 3.2GB of CCTV footage capture per hour per camera
  - Estimated use of 6 million CCTV cameras in the UK
- Semantic queries are an efficient method to facilitate human exploration
Google Image Search “Armed Robbery”
On Semantic Descriptions

Advantages
1. No (feature/ sensor) ageing
2. Available at a distance/ low resolution/ poor quality
3. Fit with human description/ forensics
4. Complement automatically-perceived measures
5. Need for search mechanisms

Disadvantages
1. Psychology/ perception
2. Need for labelling
Semantically-Mediated Information Fusion

- **Ontologies** serve as mediators for information fusion
- Computation and association of **certainty coefficients** with information elements
- Representation of trust and **provenance** information – critical to assessments if information quality and reliability
- **Fusion processes** exploit semantic axioms
  - type recognition
  - identity inference
  - identification of potential **conflicts** / inconsistencies
Semantically-Mediated Military Data Fusion

Information Sources

Feature extraction

Feature analysis

Ontologies

Fusion Processes

Classifications

Sensor data

Context

Weather

Position

Threat

Interoperability

Guo, Damarla, Nixon et al., Milcom 2007
Semantically-Mediated Biometric Fusion

Information Sources

Feature extraction
Feature analysis

Ontologies

Appearance
Sex
Ethnicity
Build
Luggage

Sensor data

Fusion Processes

Classifications

Samangooei and Nixon, IEEE BTAS 2008
Illustration

Feature space

Measure 1

Measure 2

Decision boundary

Two different classes

Initial position, two sensors compromised

After semantically-enhanced fusion

poor sensor

good sensor
Common themes

1. Feature set potency
2. Quality of feature extraction and description
3. Application scenarios
4. Need to learn semantic labels
Related work

- Feature extraction
- Feature set selection
- Classification
- Fusion
- Soft biometrics (soft = cannot be used to discriminate within populations, but between them)
- Vehicle recognition
Gait biometrics

As a biometric, gait is available at a distance when other biometrics are obscured or at too low resolution.

By computer vision, it needs moving feature extraction.

Nixon, Chellappa and Tan, Human ID Based on Gait, Springer 2005
Velocity moments

- Extension of spatial moments
- Applied to silhouettes
- Selected by ANOVA

\[ A_{mn\mu\gamma} = \sum_{\pi} \sum_{i} \sum_{j} \pi \cdot P_{i,x,y} \]
Gait recognition

- 3 moments for visualisation; subjects are clusters of 4
Translation to Real World:
Covariate Analysis

www.gait.ecs.soton.ac.uk

Shutler et al, RASC 2002
Results: Covariate Analysis

Vertex based approach vs. silhouette-based approach
Covariates: Footwear, Clothing, Time

Results: Covariate Analysis

Footwear

Clothing

Load Carriage

Walking Speed
Viewpoint invariant recognition

Can *compensate* for camera view, 25°-144°

Various *assumptions* on walking

Recognition immune to *view* and *covariates*
Does it really work?
Recent Conviction

Bag snatcher case, London 2008
Exploring Human Descriptions

- We explore semantic descriptions of:
  - physical traits
  - semantic terms
  - visible at a distance

Samangooei and Nixon, SAMT 2008
Samangooei and Nixon, IEEE BTAS 2008
Terms and Traits

• What traits are described and what terms used depends on situation
  – Mug shots vs CCTV
• Traits chosen such that
  – They are visible at a distance
  – Mentioned consistently
• Complementary qualitative terms selected
  – To avoid issues with value judgments
Traits and Terms

Global Features
- Features mentioned most often in witness statements
- Sex and age quite simple
- Ethnicity
  - Notoriously unstable
  - There could be anywhere between 3 and 100 ethnic groups
  - We’ve chosen 3 “main” subgroups and 2 extra to match UK Police force groupings

Subcategories:
- Global
  - Sex
  - Ethnicity
  - Skin Colour
  - Age
- Body Shape
  - Figure
  - Weight
  - Muscle Build
  - Height
  - Proportions
  - Shoulder Shape
  - Chest Size
  - Hip size
  - Leg/Arm Length
  - Leg/Arm Thickness
- Head
  - Hair Colour
  - Hair Length
  - Facial Hair Colour/Length
  - Neck Length/Thickness
Traits and Terms

Body Features

- Based on whole body description stability analysis by MacLeod et al.
  - Features showing consistency by different viewers looking at the same subjects
- Mostly comprised of 5 point qualitative measures
  - (Very Thin -> Very Fat, Very Short -> Very Long)
- Most likely candidate for association with gait
Traits and Terms

Head Features
- Mentioned **consistently** by people even at **long distances**
- Prominent area of **gaze**
- **Hair Length** and **colour** inherently connected with style
  - Many different hair **styles**
  - **Avoided** due to unfamiliarity of annotators

Body Shape
- **Global**
  - Sex
  - Ethnicity
  - Skin Colour
  - Age
- **Body Shape**
  - Figure
  - Weight
  - Muscle Build
  - Height
  - Proportions
  - Shoulder Shape
  - Chest Size
  - Hip size
  - Leg/Arm Length
  - Leg/Arm Thickness

Head
- **Hair Colour**
- **Hair Length**
- **Facial Hair Colour/Length**
- **Neck Length/Thickness**
Annotation Interface

Web interface constructed to gather annotations against any source

Designed to deal with issues of human description/perception

**Memory issues:** view a subject as many times as required

**Defaulting:** explicitly asked to fill out every feature

**Value Judgments:** categorical qualitative values.

**Subjective variables:** collect description of annotators
Adding semantic labels
Automatic Gait Signatures

• Gait biometrics work under the constraints of CCTV
  – Long distance to camera, noisy data etc.
• Several gait signatures can be generated from video
  – Statistical vs model based
• Average Silhouette baseline algorithm chosen for these preliminary tests
• We use the Southampton dataset
  – 115 subjects
  – At least 6 fronto-parallel videos of natural gait
Average Silhouettes Signature

- Background is taken from each frame and pixels thresholded resulting in a binary image
- Normalise silhouettes by height to account for distance
- Add all silhouettes together and divide by the number of frames
- Resulting image is the signature

Veres, Carter and Nixon, CVPR 2004
Average Silhouettes Signature

• Background is taken from each frame and pixels thresholded resulting in a binary image
• Normalise silhouettes by height to account for distance
• Add all silhouettes together and divide by the number of frames
• Resulting image is the signature
Average Silhouettes Signature

- Background is taken from each frame and pixels thresholded resulting in a binary image
- Normalise silhouettes by height to account for distance
- Add all silhouettes together and divide by the number of frames
- Resulting image is the signature
Recognition Analysis

- Generate **average** gait silhouette
- **Adjoin** semantic description
- Perform **recognition**
Biometrics: recognition capability

![Graphs showing performance metrics for different feature sets]

- Semantic features
- Semantics, fused
- Resampled semantics
- Automatic features

Samangooei and Nixon, IEEE BTAS 2008
Retrieval Analysis

• Construct *semantic space* from the training set
• *Project* test set into the space with semantic features set to zero
• Construct a *document* per semantic annotation and project into the space
• Order subjects by *cosine* distance to query
Latent Semantic Analysis

• We use Latent Semantic Analysis to facilitate content based retrieval
• This process involves formulating observations of subjects as a matrix of
  – Documents (subjects) with
  – Terms (annotations and average silhouette signatures)
• Based on observations, we construct a linear algebraic semantic space
  – The axis of which are the eigenvectors of the co-occurrence matrices
  – We ignore axis with lower eigenvalues, those likely to represent noise
  – (related to PCA)
• In this space similar geometric position implies similar meaning
Biometrics: perspicacity of labels

Samangoeei and Nixon, SAMT 2008
Biometrics: retrieval by labels

Query: Male

Query: Adolescent

Samangooei and Nixon, SAMT 2008
Successful Results

- **Hair Length** (Long vs Short)
Successful Results

- Age (Pre-Adolescent vs Young Adult)
Failed Results

- **Skin Colour** (Black vs White vs Tanned)
Vehicle Recognition Schema

Guo, Nixon and Damarla,
Information Fusion 2008
Acoustic Sensor Data

Acoustic Sensors

Sensors array configuration

Test track and arrays’ positions
Vehicle Sounds

Spectra

Harmonics

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Damarla and Phipps, Proc. SPIE, 5796, 2005
Recognition by fusion
Confusion matrix

Tracked \( t \); Wheeled \( w \)

<table>
<thead>
<tr>
<th></th>
<th>V1(_t)</th>
<th>V2(_t)</th>
<th>V3(_w)</th>
<th>V4(_w)</th>
<th>V5(_w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1(_t)</td>
<td>768</td>
<td>50</td>
<td>20</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td>V2(_t)</td>
<td>72</td>
<td>1592</td>
<td>134</td>
<td>196</td>
<td>121</td>
</tr>
<tr>
<td>V3(_w)</td>
<td>47</td>
<td>126</td>
<td>1805</td>
<td>341</td>
<td>258</td>
</tr>
<tr>
<td>V4(_w)</td>
<td>21</td>
<td>91</td>
<td>141</td>
<td>825</td>
<td>101</td>
</tr>
<tr>
<td>V5(_w)</td>
<td>3</td>
<td>74</td>
<td>119</td>
<td>119</td>
<td>1016</td>
</tr>
</tbody>
</table>
Possible Ontologies

- Sensor(s)
- Sequences of events
- Data
- Support
Semantically mediated fusion

Training Data

Parameter estimation

Semantic annotation

Feature extraction
- Harmonics
- Key frequencies

Test Data

Information fusion

Mediation
### Accuracy

1. **Annotation accuracy**

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Strong</th>
<th>Weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation accuracy</td>
<td>96.5</td>
<td>87.4</td>
</tr>
</tbody>
</table>

2. **Classification accuracy (semantic enrichment)**

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Direct</th>
<th>Ontology-based weaker</th>
<th>Ontology-based stronger</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>73.4</td>
<td>71.9</td>
<td>79.1</td>
</tr>
</tbody>
</table>

3. **Classification accuracy (semantic-med fusion)**

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Direct fusion</th>
<th>Semantically-mediated fusion</th>
<th>Ontology-based fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weaker</td>
<td>83.3</td>
<td>84.23</td>
<td>77.4</td>
</tr>
<tr>
<td>Stronger</td>
<td>83.3</td>
<td>84.18</td>
<td>85.3</td>
</tr>
</tbody>
</table>

---
Conclusions (and where does this take us?)

- Can improve performance by including semantic attributes
- Can learn semantic annotations and enrich classification
- Need more data and more labels
- Exploit the link between feature extraction and selection, fusion and semantic descriptions

- .....questions?
Some papers


+ journal versions submitted