Labannotation
A Language for InforMotion
Introduction

“Body language is a powerful source of information about human emotions and intentions” [1]

Dr. Beatrice de Gelder, 2007

[1] Beatrice de Gelder. Factsheet. C0mmunication with Emotional B0dy Language (COBOL), EU FP6
Problem Context

Artificial System
Non-Vocal Language Model

What to Interpret?  How to Interpret?  Where to apply?

Abstract Sensing
Non-Vocal Language (human motion)

Environment

Photograph: the smart room setup @ Mobile Robotics Laboratory, Institute of Systems and Robotics, University of Coimbra, Portugal.
Problem Context

- Automatic Recognition and Annotation

Photograph: the smart room setup @ Mobile Robotics Laboratory, Institute of Systems and Robotics, University of Coimbra, Portugal.
Problem Statement

What?
- Extract heterogeneous information from body language
- using an abstract representation of human activity,
- via a scalable + flexible + accurate model
- for Automatic activity recognition and characterization.
Proposed Solution

How?
- Using an abstract concept of motion,
- we developed a Hierarchic Model
- considering input abstraction.
- Towards an automatic recognition and annotation tool,
- each layer supports different modelling methodologies
- Given those methodologies capability of being represented as Dynamic Bayesian Networks.
Hardware Support

- **Acquisition Devices**
  - Synchronized Image + Polhemus or Image + Xsens
  - Default Sampling: **40Hz**
  - Target output: Cartesian trajectories

<table>
<thead>
<tr>
<th>Device</th>
<th>Description</th>
<th>Sampling Frequency</th>
<th>Image</th>
<th>Output</th>
</tr>
</thead>
</table>
| Polhemus Liberty| Magnetic Sensor              | 120Hz              | [Image](#) Polhemus Liberty| • x,y,z
                                                                            | • Roll, Pitch, Yaw           |
| MVN Suit Xsens  | Inertial Measuring Unit      | 240Hz              | [Image](#) MVN Suit Xsens  | • Accelerometer
                                                                            | • Gyroscope
                                                                            | • Inferred x,y,z             |
| Video Camera    | Tracking Algorithm           | 40Hz               | [Image](#) Video Camera    | • Tracked x,y               |
The artificial system and models should be able to:

- deal with uncertainty;
- be adaptive and flexible;
- allow for new knowledge to be easily added;
- support interpretation abstraction;
- support highly complex data (Multi-modal data);
- allow prediction of future behaviour;
Modelling Methodologies

• Unable to deal with uncertainty.

• Fail on novel inputs and complex multimodal data.

• Does not allow prediction.
• Focus on data’s intrinsic structure.

Bayesian Models

- Deterministic
- Discriminative
- Descriptive Stochastic
- Generative Stochastic
Bayesian Approach

- Challenge of generative learning: identify variables and the way they relate.
- Developing a Bayesian model has an intuitive formalism based on Bayesian Programming [2].
- Bayesian Programming defines 4 Key steps:
  1. Variables definition
  2. Decomposition
  3. Parametric Forms
  4. Question (Inference)

Bayesian models are often represented by probabilistic graphic models, where:

- Nodes represent variables
- Arrows represent conditional dependencies

Complete models of variables and their relationships, can be used to answer probabilistic queries about them, i.e. find updated knowledge of the state of a variable when other variables are observed.

For a complete specification of the Bayesian network, it is necessary to represent for each node, a probability distribution of that node conditional to its dependencies.

\[ P(\text{node}|\text{dependencies}) \]
There are numerous methodologies based on Bayes theorem

which have an equivalent representation as Dynamic Bayesian Networks (DBN) [3]

Bayesian Programming 3/7

Definitions: Bayes theorem

- establishes a degree of belief over a variable state

**Graphic Model**

- Prior: the initial degree of belief for A.
- Likelihood: represents previous knowledge
- Posterior: the degree of belief in A given the knowledge of B
- Normalization factor: is usually omitted for simplification.

\[
P(A | B) = \frac{P(A) \cdot P(B | A)}{P(B)}
\]
Bayesian Programming 4/7

Bayesian Approach

Which activity? (Laban^Features)

Decision via Inference (Bayes Rule) to answer

\[ P(\text{Activity} | \text{Laban } \land \text{Features}) \]

Prior Knowledge

Learning by Experimental data

Enviroment

Artificial System
Bayesian Programming 5/7

Definitions: Bayesian Inference using Bayes Rule

- Derives posterior probability based on two antecedents:
  - Prior probability
  - Likelihood function
  which derive from a probability model for the observed data.
- Inference according to Bayes theorem yields the probability of an hypothesis given observed evidence:

\[
P(\text{Hypothesis} \mid \text{Evidence}) = \frac{P(\text{Evidence} \mid \text{Hypothesis})}{P(\text{Evidence})} P(\text{Hypothesis})
\]
In a Bayesian approach, probability emerges as an alternative to logical systems.

Incompleteness

- Preliminary Knowledge
  + Experimental Data
  = Probabilistic Representation

Learning

Entropy Principles

Uncertainty

Bayesian Inference

Decision

\[ P(a \land b) = p(a)p(b | a) = p(b)p(a | b) \]
Key steps in Bayesian programming [2] summarized:

1. **Variables Definition**: identification of variables that relate to observable data and the system parameters the model aims to estimate.

2. **Decomposition**: represent each variable as probabilistic distributions according to the defined dependencies.

3. **Parameterization**: decide which type of distribution is suitable for representing/model a variable probability (Gaussian, Stochastic Matrix, Uniform,...)

4. **Question**: define the model queries, i.e. what we want to know based on what we can observe.

Step 1: Variable Definition 1/3

- Variables are defined based on Laban Movement Analysis

<table>
<thead>
<tr>
<th>Language</th>
<th>Laban Movement Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups</td>
<td>Non-Kinematic</td>
</tr>
<tr>
<td></td>
<td>Kinematic</td>
</tr>
<tr>
<td>Components</td>
<td>Effort</td>
</tr>
<tr>
<td>Qualities</td>
<td>Time</td>
</tr>
</tbody>
</table>

- Studies qualitative aspects and expressiveness of motion.
- Intentional process of patterned and orderly changes.
- Defines basic, irreducible basic motion elements.
- Motion is better studied if divided in multiple levels.
Step 1: Variable Definition 2/3

- Information hierarchy paradigm (example)
  - Variable = {..., state, ...}
### Step 1: Variable Definition 3/3

States are defined based on theories of Laban and related studies on activity description using Labanotation[4]

<table>
<thead>
<tr>
<th>Abstraction</th>
<th>Level</th>
<th>Variable</th>
<th>Designation</th>
<th>States</th>
</tr>
</thead>
<tbody>
<tr>
<td>InforMotion</td>
<td>Activity</td>
<td>$G^t$</td>
<td>Gesture</td>
<td>Bye, Punch, Point, Lift, Write</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$E_t$</td>
<td>Effort Time</td>
<td>Sudden, Sustained</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$E_{s}^t$</td>
<td>Effort Space</td>
<td>Direct, Indirect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$E_{w}^t$</td>
<td>Effort Weight</td>
<td>Free, Careful</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$E_{f}^t$</td>
<td>Effort Flow</td>
<td>Strong, Light</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$S_{f}^t$</td>
<td>Shape Flow</td>
<td>Spreading, Enclosing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$S_{sh}^t$</td>
<td>Shape Shape</td>
<td>Rising, Sinking</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$S_{sp}^t$</td>
<td>Shape Space</td>
<td>Reaching, Retreating</td>
</tr>
<tr>
<td>Mathematic</td>
<td>Feature</td>
<td>$F_{m}^t$</td>
<td>$m^{th}$ Low Level Feature</td>
<td>Class $a$, $a \in [1,k] \rightarrow \mathbb{N}$</td>
</tr>
</tbody>
</table>

Step 2: Decomposition 1/3

- Formulation Properties
  - Hierarchic
  - Scalable
  - Flexible

- Joint Distribution: \( P(\text{Activity} \land \text{Laban} \land \text{Feature}) \)

- According to Bayes rule, start the decomposition for the lowest two layers by defining:
  - a prior \( P(\text{Laban}) \)
  - and a likelihood \( P(\text{Feature}|\text{Laban}) \)
  - to compute a posterior probability (\( \propto \) inference):

\[
P(\text{Laban}|\text{Feature}) \propto P(\text{Laban})P(\text{Feature}|\text{Laban})
\]
Step 2: Decomposition 2/3

However, the model has multiple layers

- Laban and Activity are dependent.
- Hence, and according to Bayesian formulation $P(\text{Laban})$ is replaced by $P(\text{Laban} | \text{Activity})$
- and a new prior $P(\text{Activity})$ for the system state is required
- leading to a posterior decomposition yielding

$$P(\text{Laban} \land \text{Activity} | \text{Feature}) \propto P(\text{Feature} | \text{Laban}) P(\text{Laban} | \text{Activity}) P(\text{Activity})$$

Note: The model for $P(\text{Laban} | \text{Activity})$ is the Dynamic Bayesian Network representation of a Markov Model, therefore the generic Bayesian Rule holds.
Step 2: Decomposition 3/3

Expanded Decomposition of the Motion Model

\[ P(G^t) \times P(ET^t \mid G^t) \times P(ES^t \mid G^t) \]
\[ \times P(EF^t \mid G^t) \times P(SS^t \mid G^t) \]
\[ \times P(SSh^t \mid G^t) \]
\[ \times P(F_1^t \mid ET^t) \]
\[ \times P(F_m^t \mid ET^t) \]
\[ \times P(F_m^t \mid SSh^t) \]
Defining **Observation Data**

- From multiple feature generation algorithms we selected Kharunen-Loëve Transform (KLT)\[5\] to segment our data\[6\].

\[
X_n = [x_{t-w}, \ldots, x_t] \\
\frac{1}{w} \sum_{n=1}^{w} x_n \\
\frac{1}{w}(X_n - \mu)(X_n - \mu)^T \\
C \\
solve \ V^{-1}CV = D \\
\lambda_n = \text{Eigenvalues} \\
v_n = \text{Eigenvectors}
\]

\[F^t = \text{KLT} (X_n) = \{f_1^t, \ldots, f_m^t\}\]

- \(f_m^t = c\), \(c \in [1, k]\) \(\rightarrow \mathbb{N}\) and \(k = \text{number of classes}\)
- Classes c are defined dividing both ranges \([\lambda_{\text{min}}, \lambda_{\text{max}}]\) and \([v_{\text{min}}, v_{\text{max}}]\) in \(k\) equidistant intervals.

---


\[6\] Luis Santos and Jorge Dias. Motion patterns: Signal interpretation towards the Laban Movement Analysis semantics. In Technological 466 Innovation for Sustainability: IFIP Advances in Information and Communication Technology, 2011
Step 3: Parametric Forms 2/3

- In **Laban Layer** we define a model for estimating Laban components based on observed data features.

- The parametric distribution types of our model are:
  - Prior $P(L_n^t) = \text{uniform} \; ; \; t=0$
  - Prior $P(L_n^t) = P(L_{n}^{t-1}) \; ; \; t \neq 0$
  - Likelihood $P(F_m^t|L_n^t) = \text{stochastic matrix}$
  - A posterior defined as

$$P(L_n^t \mid F_1^t \land \ldots \land F_m^t) \propto P(L_n^t) \sum_{i=1:m} P(F_i^t \mid L_n^t)$$

Likelihood is defined upon **Experimental data (learning)**
Step 3: Parametric Forms 3/3

- In **Activity Layer** we define a model for Gesture variable as sequences of Laban states.

- Via equivalent representation, we parameterize the HMM as its equivalent Bayesian Network:
  - Prior $P(G^t) = \text{uniform} \quad ; t=0$
  - Prior $P(G^t) = P(G^{t-1}) \quad ; t \neq 0$
  - Prior $g^0 = P(G^0)$, the initial process state.
  - Likelihood $P(G^t=j | G^{t-1}=i) = A^t(i,j)$: transition stochastic matrix
  - Likelihood $P(L_n^t = j | G^t = i) = L^t(i,j)$: observation stochastic matrix
  - Posterior $= P(G^t = q_i \land G^{t+1} = q_j \mid A^t L^t g^0) =$

$$P(G^t) \times P(G^t \mid G^{t-1}) \times P(L_n^t \mid G^t)$$
Step 3: Learning 1/2

Signal → Acquisition → Processing → Annotation → Output

- Feature Generation
  - Stochastic Matrix
  - Gaussian
  - Poisson

Learned Probabilistic Distributions
Step 3: Learning 2/2

- Consider variables $Feature = \{1, \ldots, 20\}$ and $Laban = \{0, 1\}$.
- A stochastic matrix $M$ represents $P(Feature=i|Laban=j)$

$$M(i, j) = \frac{m_{i,j}}{\sum_{q=1}^{j} m_{iq}}, \text{ where } \sum_{i} m_{i,j} = 1$$

- Example:

- Remaining stochastic matrices are defined analogously.
Step 4: Question 1/3

- Given the previous formulation, we are now able to query our model for information.
- The global question is given by the posterior distribution for the complete model:

\[
P(G^t | G^{t-1} \land L_n^t \land F_m^t)
\]
Step 4: Question 2/3

- Apart from the global question, the model allows to ask intermediate questions.
- Parametric questions:
  - allowing access the different information levels
- Which Laban state given the observed trajectory.
  
  \[ P(L_n^t \mid F_m^t) \]

- and which gesture given the previous estimated gesture and the knowledge of the current Laban state
  
  \[ P(G^t \mid L_n^t \land G^{t-1}) \]
Step 4: Question 3/3

- We answer the questions asked to a model via Inference.
- There are numerous inference methods, from which we use one called Maximum A Posteriori (MAP) base on Bayes formulation.

- Inference using MAP, generically maximizes Bayes rule as:

\[
\theta_{MAP}(x) = \arg \max_{\theta} P(x | \theta)P(\theta)
\]
Performance

Upon the mathematical definition of the model, we aim to test its classification performance:

- Evaluate noise, feature selection and error propagation
- $10^4$ simulation tests (statistical significance)
- Performance metrics: convergence speed and accuracy
  - Speed is measured in model iterations
  - $L$ converges to correct class $q$ with $P(L=q) \geq 0.999$
Noise Performance

- **Question:** $P(\text{Component State} | \text{Observed Feature})$
- Defined experimental threshold: $\sigma_{\text{noise}} \approx 3\sigma_{\text{likelihood}}$
- 100% Accurate for $\sigma_{\text{noise}} \leq 2\sigma_{\text{likelihood}}$
- Average convergence time $\approx 10$ iterations
Feature Selection Performance

- **Question:** $P(\text{Component State} \mid \text{Observed Features})$

- **Good Feature Selection:**
  - 100% Accurate
  - Convergence $T \propto (\text{number of features})^{-1}$

- **Poor Feature Selection:**
  - 100% Accurate for $\sigma_{\text{noise}} \leq 2\sigma_{\text{likelihood}}$
  - Accuracy $\propto (\text{number of features})^{-1}$

![Graphs showing accuracy and iterations vs. number of observations]
Hierarchic Topology Performance

- Hierarchic vs. non-hierarchic
- Comparison by removing layer L
- Scenarios based on node connectivity
- Assumption: connectivity affects feature quality ($\sigma$).
Hierarchic Topology Performance

- **Biased Signal**
  - $T \approx 5$ iterations
  - 100% Accurate

- **Unbiased Signal**
  - $T \approx 20$ iterations
  - Accuracy $\geq 95$

- Accuracy is virtually unaffected.
- Allows extracting more information
Performance Results Summary

- The model withstands considerable noise
  - Experimental threshold ($\sigma_{\text{noise}} \approx 3\sigma_{\text{likelihood}}$)
  - Average convergence in 10 iterations

- A careful feature selection should be made.

- When features are not properly selected:
  - Accuracy $\propto$ (number of features)$^{-1}$

- Hierarchy does not affect accuracy significantly ($\approx 100\%$).
  - and it allows extracting more information.
Real data experiments

- To validate the model, we conducted experimental results with real motion data.
- Ground truth tests with 3-D high resolution data
- Validation on 2-D trajectories extracted from images:
  - MRL dataset
  - KTH dataset
  - Weizmann dataset
- Classification confusion tables for both inforMotion levels
- Benchmarking
The Mobile Robotics Laboratory (MRL) database has 93 gesture sequences.

It encompasses 5 different gestures performed by 5 different performers.

The data is mainly stored pair wise:
- Images + Polhemus
- Images + MVN suit
- Older branches of the database may be stored in Polhemus data only.

The data is synchronized and stored in XML format.

The DB is being processed for public release.
MRL Dataset

- 5 Different Gestures
  - Bye
  - Lift
  - Point
  - Punch
  - Write
- High resolution + Image data
- ≈100 activity performances

- The Database is currently being reformulated for public release, and datasets are available upon request.
Results – Laban Layer Classification

- **MRL Dataset**
  Compare classified state with ground truth annotation:
  \[ T.P.\% = \frac{\text{correct classified}}{\text{total classified}} \times 100 \]

- **Average Accuracy 2D:**
  - 82.98%

- **Average Accuracy 3D:**
  - 80.94%
Results – 3-D Data (Proof of concept)

- MRL Dataset

Confusion matrix for $P(G^t|L_n^{t^F_m})$ – Activity Layer

<table>
<thead>
<tr>
<th></th>
<th>Lift</th>
<th>Write</th>
<th>Punch</th>
<th>Point</th>
<th>Bye</th>
<th>T.P.%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lift</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100,00</td>
</tr>
<tr>
<td>Write</td>
<td></td>
<td>18</td>
<td></td>
<td>1</td>
<td></td>
<td>94,74</td>
</tr>
<tr>
<td>Punch</td>
<td></td>
<td></td>
<td>12</td>
<td>1</td>
<td></td>
<td>92,31</td>
</tr>
<tr>
<td>Point</td>
<td>1</td>
<td></td>
<td>1</td>
<td>26</td>
<td></td>
<td>92,86</td>
</tr>
<tr>
<td>Bye</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>18</td>
<td>94,74</td>
</tr>
</tbody>
</table>

94,62
# Results – 2-D Data

## MRL Dataset

Confusion matrix for $P(G^t|L_n^{t^F_m})$ – Activity Layer

<table>
<thead>
<tr>
<th></th>
<th>Lift</th>
<th>Write</th>
<th>Punch</th>
<th>Point</th>
<th>Bye</th>
<th>T.P.%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lift</td>
<td>13</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>92.86</td>
</tr>
<tr>
<td>Write</td>
<td></td>
<td>18</td>
<td></td>
<td></td>
<td>1</td>
<td>94.74</td>
</tr>
<tr>
<td>Punch</td>
<td></td>
<td></td>
<td>11</td>
<td>2</td>
<td></td>
<td>84.62</td>
</tr>
<tr>
<td>Point</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>25</td>
<td>89.29</td>
</tr>
<tr>
<td>Bye</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17</td>
<td>89.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>90.32</strong></td>
</tr>
</tbody>
</table>
## Results – 2-D Data

### KTH Dataset

<table>
<thead>
<tr>
<th></th>
<th>Boxing</th>
<th>Clap</th>
<th>Waving</th>
<th>Run</th>
<th>Walk</th>
<th>T.P.%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boxing</td>
<td>61</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td>81.30</td>
</tr>
<tr>
<td>Clap</td>
<td></td>
<td>72</td>
<td>3</td>
<td></td>
<td></td>
<td>96.00</td>
</tr>
<tr>
<td>Waving</td>
<td>12</td>
<td></td>
<td>63</td>
<td></td>
<td></td>
<td>84.00</td>
</tr>
<tr>
<td>Run</td>
<td></td>
<td></td>
<td></td>
<td>72</td>
<td>3</td>
<td>96.00</td>
</tr>
<tr>
<td>Walk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Confusion matrix for $P(G^t|L^n_{i}^{t}F_m^{t})$ – Activity Layer

91.50
## Results – 2-D Data

### Weizmann Dataset

Confusion matrix for $P(G^i|L_n^{i\wedge F_m^i})$ – Activity Layer

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>JJ</th>
<th>J</th>
<th>R</th>
<th>W</th>
<th>W1</th>
<th>W2</th>
<th>T.P.%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bend</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100,00</td>
</tr>
<tr>
<td>Jump Jack</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100,00</td>
</tr>
<tr>
<td>Jump</td>
<td></td>
<td>3</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>66,67</td>
</tr>
<tr>
<td>Running</td>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td>100,00</td>
</tr>
<tr>
<td>Walking</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
<td>88,89</td>
</tr>
<tr>
<td>Wave (1 hand)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>88,89</td>
</tr>
<tr>
<td>Wave (2 hands)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td></td>
<td>100,00</td>
</tr>
</tbody>
</table>

92,06
Experimental Video

Laban and Gesture Classifier Demo
V1.0

Summary:
- Gesture: Punch
- Actor: Kamrad
- Trial Number: 3
- Duration: 7.8 seconds

The source code for the classifier is partially implemented in ProSTv2.1 (www.probeyes.com)
## Benchmark

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Year/Reference</th>
<th>T.P.%</th>
<th>Classifying Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRL</td>
<td>2011 / [7]</td>
<td>67,16</td>
<td>Bayesian Network + Gaussian Likelihoods</td>
</tr>
<tr>
<td></td>
<td>InforMotion</td>
<td>90,32</td>
<td>Dynamic Bayesian Network + 2-D Data</td>
</tr>
<tr>
<td></td>
<td>InforMotion</td>
<td>94,62</td>
<td>Dynamic Bayesian Network + 3-D Data</td>
</tr>
<tr>
<td>KTH</td>
<td>2009 / [8]</td>
<td>90,00</td>
<td>K-Nearest Neighbour</td>
</tr>
<tr>
<td></td>
<td>InforMotion</td>
<td>91,50</td>
<td>Dynamic Bayesian Network + 2-D Data</td>
</tr>
<tr>
<td></td>
<td>2009 / [9]</td>
<td>93,30</td>
<td>Efficient Nearest Neighbour</td>
</tr>
<tr>
<td></td>
<td>2011 / [10]</td>
<td>93,98</td>
<td>Native Bayes Mutual Information Maximization</td>
</tr>
<tr>
<td></td>
<td>2011 / [13]</td>
<td>90,00</td>
<td>Cluster Transition Maps</td>
</tr>
<tr>
<td></td>
<td>InforMotion</td>
<td>92,06</td>
<td>Dynamic Bayesian Network + 2-D Data</td>
</tr>
<tr>
<td></td>
<td>2010 / [14]</td>
<td>93,50</td>
<td>K-Nearest Neighbour</td>
</tr>
<tr>
<td></td>
<td>2010 / [15]</td>
<td>98,60</td>
<td>Nearest Neighbour</td>
</tr>
</tbody>
</table>
InforMotion System

Comparison is made to recent works (2009/12)

Greatest improvement was observed from our previous work.

Results are within state of the art performance.

The presented Bayesian model scales to, and classifies different information while maintaining state of the art accuracy.

Photograph: the smart room setup @ Mobile Robotics Laboratory, Institute of Systems and Robotics, University of Coimbra, Portugal.
Benchmark Bibliography


[10] Hong-Bo Zhang, Shao-Zi Li, Xian-Ming Lin, and Bi-Xia Liu. The constrast between motion and appearence representation of stip in human action classification. In IEEE International Conference on Computer Science and Automation Engineering (CSAE), 2011


Conclusions

- Presented a scalable, hierarchic model for interpreting heterogeneous body language information;
- Defined a core space state based on LMA.
- Compose complex activities from basic information;
- Supports different Bayesian methodologies;
- Model performance tested under different conditions;
- Validation performed on external datasets;
- Trained Laban likelihoods usable across datasets;
- State of the art accuracy;
- Real-time performance.
Future work: Laban Applicability 1/2

What?
- We have activity represented in Laban Space.
- And want to define Motion Signatures;
- In order to identify people based on how they move.
Future work: Laban Applicability 2/2

- **Identifying personnel in restricted environments**
  - Monitored environments have capability to track persons.
  - Database of cleared access personnel
  - Observing unauthorized/non-identified person, would lead the model to non-convergent state and trigger occurrence.

- **Grouping Large Populations by Behaviour Patterns**
  - Exponentially large database: possible similar Laban output.
  - Similar Laban output → similar behaviour performance.
  - Bayesian nature allows model to deal with noise.
  - Group similar individuals as behavioural clusters.
  - Trigger responses based on identified behaviours.
Problem Statement

How?

- Human motion is influenced by unique physical and psychological characteristics.
- Define Signature vectors in Laban Space.
- Identify persons based on how they move.
- using Bayesian classifiers to develop a recognition model.
Signature Approaches

Three different signature approaches:

- **Linear**: Compute signature from R vector Pearson’s correlation coefficient $\rho$ [14] for each $r_{comp}$,
  \[
  \rho_{i,j} = \frac{\text{cov}(i,j)}{\sigma_i \sigma_j}
  \]

- **Topological**: consider each $r_{comp}$ as a topological node and compute the connectivity matrix.

- **GMM**: R is a coordinate in Laban Space. Decompose R clusters using GMM, using its parameters $\{\mu, \Sigma\}$ as signature features.

- Apply Singular Value Decomposition to signature matrices using Eigenvectors and values as features.

---

Step 1: Variables definition

- We define two different signatures variable types:
  - Approach A: considers each sub-signature $V$ as an independent variable
  - Approach B: aggregate all sub-signature vectors in a matrix, and reduce its dimension to generate a global signature $G$

<table>
<thead>
<tr>
<th>Level</th>
<th>Variable</th>
<th>Designation</th>
<th>States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person</td>
<td>ID$^t$</td>
<td>Person</td>
<td>Ricardo, Luis, Kamrad</td>
</tr>
<tr>
<td>Signature</td>
<td>$V_{\text{sig}}^t$</td>
<td>Sub-Signature</td>
<td>${v_1, ..., v_m}, v_m \in \mathbb{R}$</td>
</tr>
<tr>
<td></td>
<td>$G_{\text{sig}}^t$</td>
<td>Global Signature</td>
<td>${g_1, ..., g_n}, g_n \in \mathbb{R}$</td>
</tr>
<tr>
<td>Feature</td>
<td>$R^t$</td>
<td>Low Level Feature</td>
<td>$r_{\text{comp}} = P(L_n^t=q_0)$</td>
</tr>
</tbody>
</table>
Bayesian model: identify persons based on signatures.

Consider two different models for signatures approaches A and B.
A decomposition of a Joint distribution may include all distributions.

From the global joint distribution we make the desired queries.

Step 2: Decomposition

\[
P(ID^t \land V_n^t \land L_n^t) = \\
P(ID^t) \times P(ID^t \mid ID^{t-1}) \times P(V_n^t \mid ID^t \land L_n^t) \times P(L_n^t \mid V_n^t)
\]

\(L_n^t, V_n^t\) are vectors composed of independent var \(l_n\) and \(v_n\) respectively.
Step 3: Parametric Forms

\[
P(ID^t) = \begin{cases} 
  \text{Uniform}, & t = 0 \\
  P(ID^{t-1}), & t \neq 0
\end{cases}
\]

\[
P(ID^t | V_n^t \land L_n^t) = P(ID^t) \times P(V_n^t | I \land L_n^t) \times P(V_n^t | ID^t)
\]

\[
P(ID^t | G_n^t \land L_n^t) = P(L_n^t | G_n^t) \times P(G_n^t | ID^t) \times P(ID)
\]

All other distributions are stochastic matrices.

Step 4: Global question for approach A

\[
P(ID^t | V_n^t \land L_n^t)
\]

Step 4: Global question for approach B

\[
P(ID^t | G_n^t \land L_n^t)
\]
Experimental Setup Description

- 103 different activity sequences (27x10^3 feature vectors).
- Only 30% of samples are used for training the model.
- 3 different performers
- Synchronized images and high resolution data (120Hz)
- Proof of concept tests with 3-D high resolution data
  - MRL dataset
- Recognition classification accuracy in 4 different scenarios:
  1. Feature Selection, $P(ID^t) = \text{Uniform}, \forall t$
  2. Feature Selection, $P(ID^t) = P(ID^{t-1}), \forall t \neq 0$
  3. No Feature Selection, $P(ID^t) = \text{Uniform}, \forall t$
  4. No Feature Selection, $P(ID^t) = P(ID^{t-1}), \forall t \neq 0$
Results

- **Approach A**
  - Scenario 2, presents nearly perfect accuracy.
  - Only scenario 3 has poor results for one ID class.
  - Generated signatures are visibly different for different persons.
Results

- **Approach B**
  - Scenario 2 is the best approach.
  - Yet, there is a lot of misclassified samples.
  - Global signatures present poor discriminative capabilities, due to the applied reduction algorithm.
Conclusions

- Presented a Bayesian model for person recognition
- Developed 3 different signatures in Laban Space.
- Proved the concept in our MRL dataset.
- When a good feature selection and recursive estimation is applied, accuracy $\approx 100\%$.
- Successfully identified real application scenarios.

Future Work

- Demonstrate the model in external datasets.
- Develop a functional application.
Questions?

Thank you!

Luis Santos – luis@isr.uc.pt
jorge@isr.uc.pt

Jorge Dias – jorge.dias@kustar.ac.ae

Mobile Robotics Laboratory

Instituto de Sistemas e Robótica

Departamento de Engenharia Electrotécnica e de Computadores
Universidade de Coimbra
Pinhal de Marrocos - Polo II
3030-290 COIMBRA - Portugal