Imperial College London

CeRTH
Informatics & Telematics Institute

ITI Presentation

Activity related recognition in *AmI* environments

Mr. Anastasios Drosou

(Dr. Tzovaras' Group)

Outline

- **Introduction**
- Activity Detection
- Activity related recognition
  - Human Tracking
  - Reaching Part
  - Grasping Part
- Classifiers
- Results
- Conclusions

Definition of Biometrics

"Biometrics measure the unique physical or behavioural characteristics of individuals as a means to recognize or authenticate their identity."

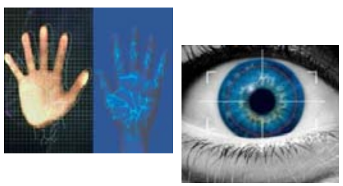
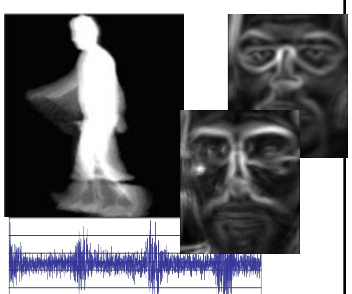
→ Establish someone's identity based on who he/she is rather than on what he/she poseses (e.g. ID cards,) or what he/she remembers (e.g. password).

- <u>Verification problem</u>: *Is the user Mr. X? 1:1 comparison.*
- <u>Identification</u>: *Who is the user? 1:many comparison.*

Imperial College London

CeRTH ITI

Types of Biometrics

	Description	Types	Pros	Cons
Physiological Biometrics 	A physical attribute unique to a person	Fingerprint, palm, iris, geometry, retina, facial characteristics, etc.	High Accuracy.	Obtrusive, Uncomfortable, Easy to spoof, Intolerant of changes over time.
Behavioural Biometrics 	Traits that are learned or acquired from a person	Facial Dynamics, Gait, Voice, Key-stroke Patterns, Activity related signals, etc.	Unobtrusive, Difficult to spoof, Continuous & on-the-move authentication, Rare changes.	Lower accuracy, yet...

Imperial College London
CeRTH ITI

Motivation

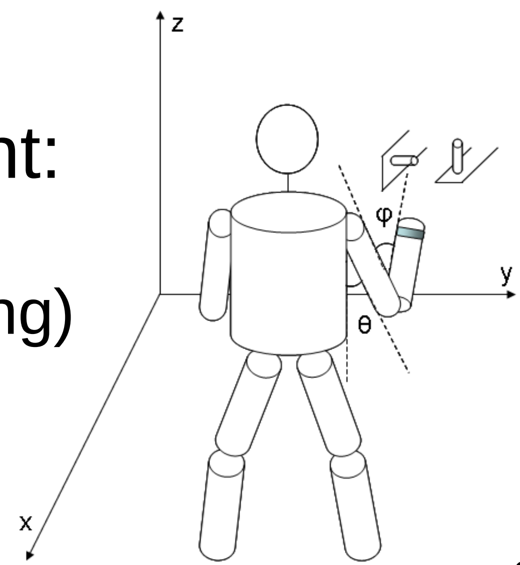
We claim **recognition potential** because:

1. Personal behavioural patterns in everyday movements (i.e. gait, grimaces, standard movements, etc)
2. Physiology of the human body (height, arm length, finger lengths, etc.)
3. Bodymetric restrictions (i.e. impairments, etc.)
4. Minimum Jerk Model (Flash & Hogan)
5. Minimum Torque Change Model (Uno, Kawato & Suzuki)
6. End state comfort effect (Rosenbaum et al.)
7. …

Imperial College London
CeRTH ITI

Case study: Prehension biometrics

- Response of the person to specific stimuli.
- Work related, everyday activities with no special protocol.
- Continuous authentication/recognition.
- Multi-activity concept for human authentication.



➢ 2 stages of a prehension movement:
   i. a fast initial movement (reaching)
   ii. a slow approach movement (grasping)

System-Level Criteria

Characteristics of a biometric that must be present in order to use the system for authentication purposes

- **Universality**
  - *Universality*: Every person should possess this characteristic
  - In practice, this may not be the case
  - Otherwise, population of non-universality must be small < 1%
- **Uniqueness**
  - *Uniqueness*: No two individuals possess the same characteristic.
  - Genotypical – Genetically linked (e.g. identical twins will have same biometric)
  - Phenotypical – Non-genetically linked, different perhaps even on same individual
  - Establishing uniqueness is difficult to prove analytically
  - May be unique, but “uniqueness” must be distinguishable
- **Permanence**
  - *Permanence*: The characteristic does not change in time, that is, it is time invariant
  - At best this is an approximation
  - Degree of permanence has a major impact on the system design and long term operation of biometrics. (e.g. enrollment, adaptive matching design, etc.)
  - Long vs. short-term stability
- **Collectability - Measurability**
  - *Collectability*: The characteristic can be quantitatively measured.
  - In practice, the biometric collection must be:
    - Non-intrusive
    - Reliable and robust
    - Cost effective for a given application
  - **Measurability:** The trait can be measured with simple technical instruments
- **Performance** – accuracy, speed, and robustness of technology used
- **Circumvention** – The trait is easily measured with minimal discomfort

….then it can potentially serve as a biometric *for a given application*.

Outline

- Introduction
- **Activity Detection**
- Activity related recognition
  - Human Tracking
  - Reaching Part
  - Grasping Part
- Classifiers
- Results
- Conclusions

Motion Templates

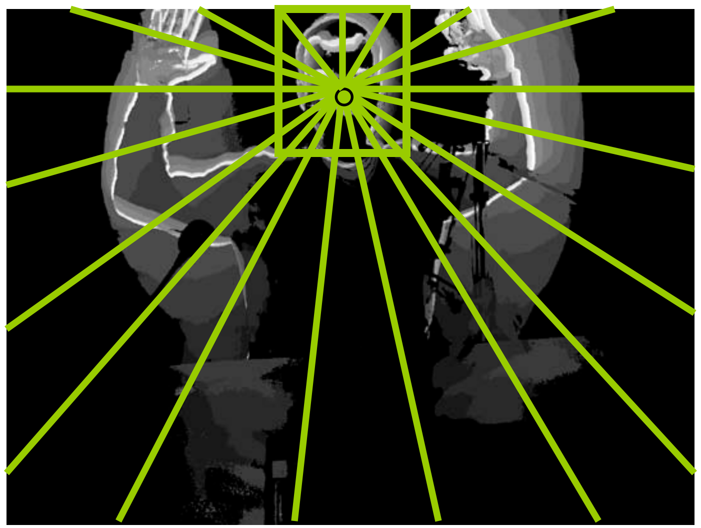
- Motion Energy Image (MEI)
  - Describes the motion energy for a given view of action.
  - Binary image.

$$E_T(x,y,t) = \bigcup_{i=0}^{\tau-1} D(x,y,t-i)$$

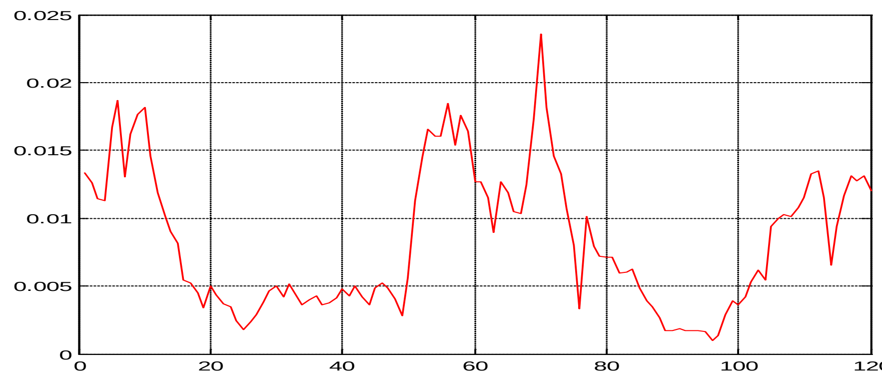
- Motion History Image (MHI)
  - Intensity image.

$$MHI_T(x,y,t) = \begin{cases} \tau & \text{if } D(x,y,t)=1 \\ max(0, MHI_T(x,y,t-1)-1) & \text{otherwise} \end{cases}$$

Radon Transforms (I)



1. Face detection
2. Centre of Face
3. RIT Extraction

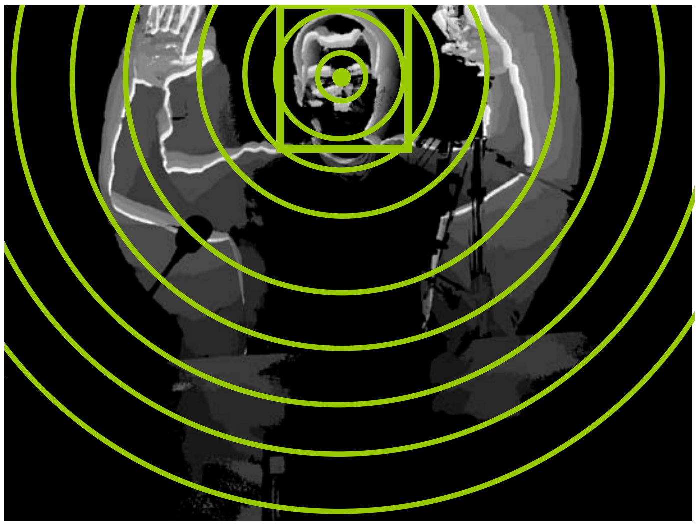


- **Radial Integration Transform (RIT):**

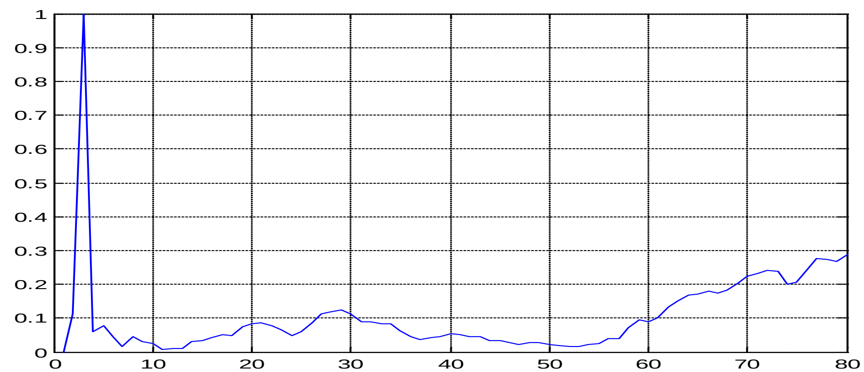
$$RIT(t\Delta\theta) = \frac{1}{J}\sum_{j=1}^{J} MHI(r_0 + j\Delta u \cos(t\Delta\theta), y_0 + j\Delta u \sin(t\Delta\theta))$$

for $t = 1,\ldots,T$ with $T = 360^o / \Delta\theta$

Radon Transforms (II)



1. Face detection
2. Centre of Face
3. CIT Extraction



- **Circular Integration Transform (CIT):**

$$CIT(t\Delta\rho) = \frac{1}{T}\sum_{t=1}^{T} MHI(x_0 + k\Delta\rho\cos(t\Delta\theta), y_0 + k\Delta\rho\sin(t\Delta\theta))$$

$$\text{for } k = 1,\ldots,K \text{ with } T = 360^o / \Delta\theta$$

Similarity Metrics

- Euclidian Distance

$$D_E(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{x}-\mathbf{y})^2} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

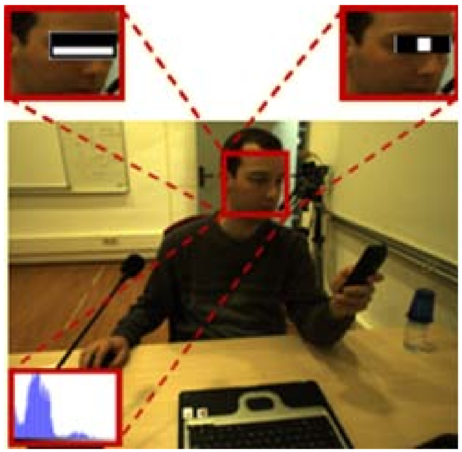
- Correlation Factor - Coefficients

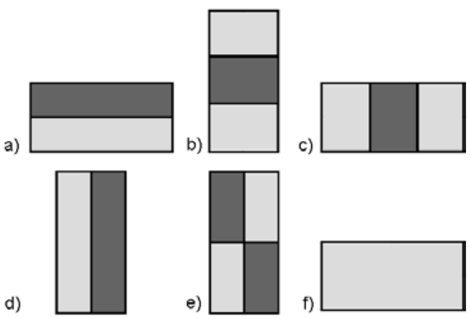
$$corr(x,y) = \rho_{\mathbf{x},\mathbf{y}} = \frac{\mathrm{cov}(\mathbf{x},\mathbf{y})}{\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}} = \frac{\mathrm{E}((\mathbf{x}-\mu_{\mathbf{x}})(\mathbf{y}-\mu_{\mathbf{y}}))}{\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}}$$

Human Body Tracking

- Camera based tracking:
  - Custom ITI Tracker
  - OpenNI Library Tracker
- *Ascension Technology Corp.* Magnetic Tracker (ground truth).
- *Cyberglove*® finger tracking.

Face Detection & Tracking



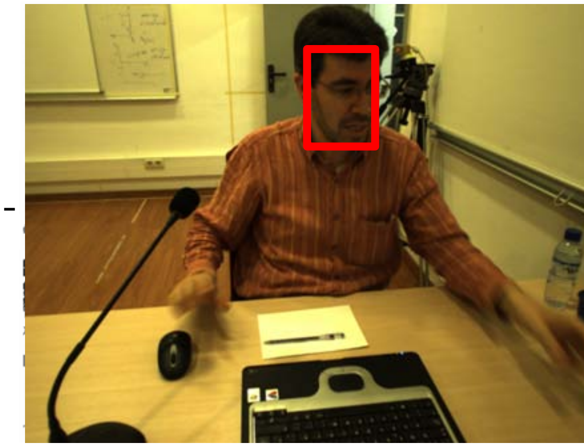


- Viola & Jones's real-time face detection method:
  - *T* weak classifiers form a strong one,
  - integral Image,
  - haar like features (subtracting the pixel values in the dark from the bright rectangles),
  - AdaBoost algorithm.

- Mean Shift algorithm:
  - Bases on spatial and colour information between sequential frames.
  - Bhattacharyya distance

Background Extraction – Skin Colour Filtering – Motion Detection

- Location of the face
- Disparity images (stereoscopic Camera)

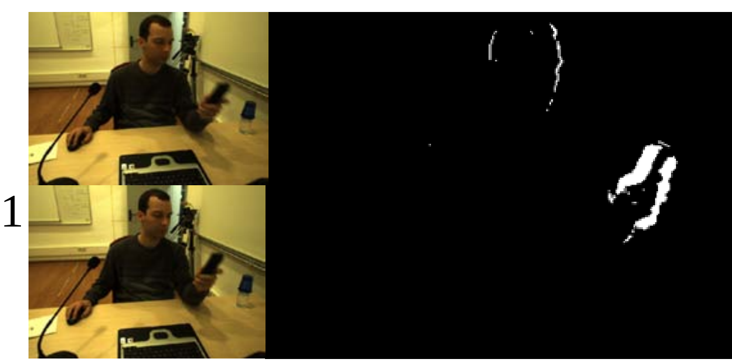


$$\boxed{D(I) = S(I) \cap B(I)}$$

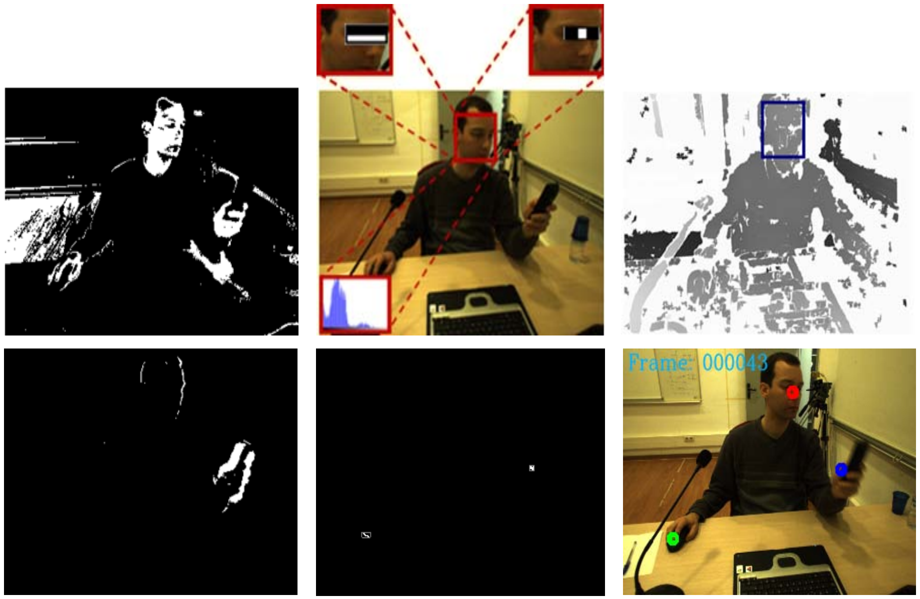
- No training required.
- Computational inexpensive.
- Simple approach - explicit rules.
- Skin cluster boundaries of RGB & HSV.

$$\boxed{M_t(I) \equiv D_1(I) - D_2(I)}$$

$$MHI_t(x,y) = \begin{cases} 2, & \text{if } M_t(I(x,y)) = 1 \\ max(0, MHI_{t-1}(x,y) - 1), & \text{otherwise} \end{cases}$$



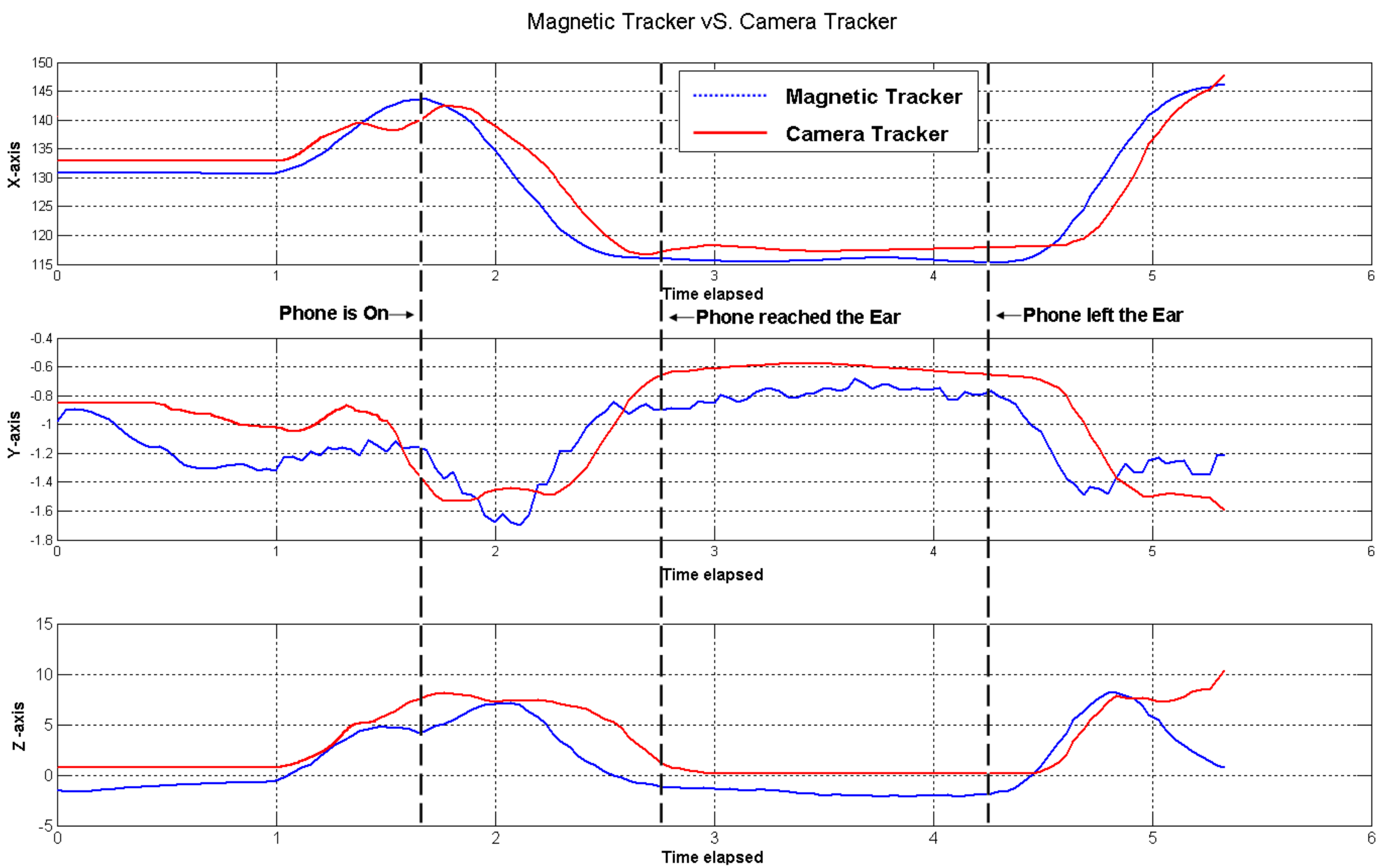
Movement's Tracking



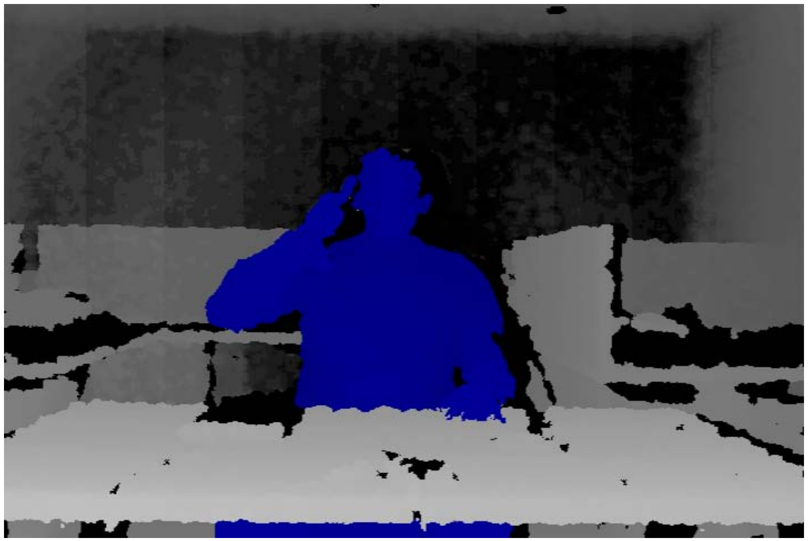
Imperial College London

CeRTH ITI

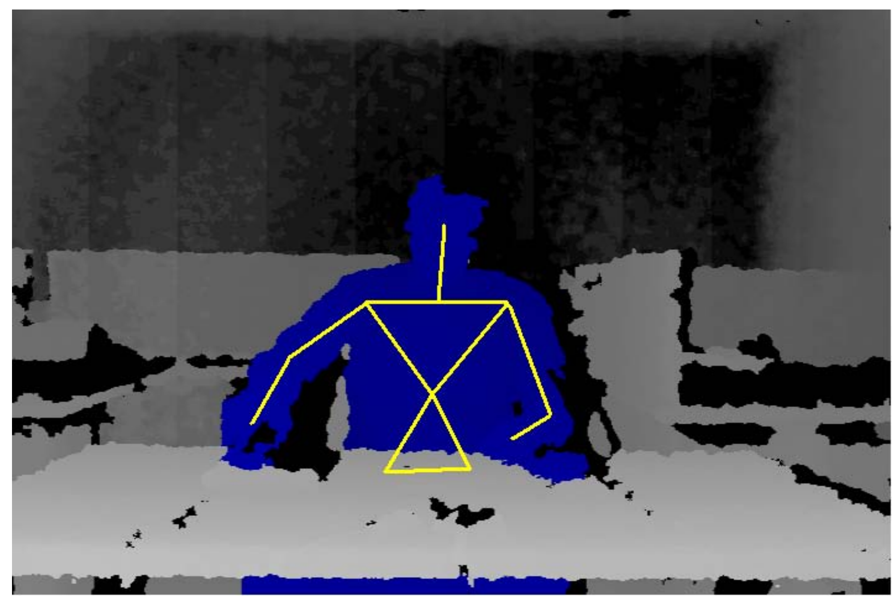
ITI Body Tracker Evaluation



Primesense® TOF Sensor - OpenNI Open Source Library



- Real time & accurate.
- Human body form recognition and segmentation ecognizes the human
- Human body tracking by simultaneousl tracking of 48 essential points of the human body in the 3D space.

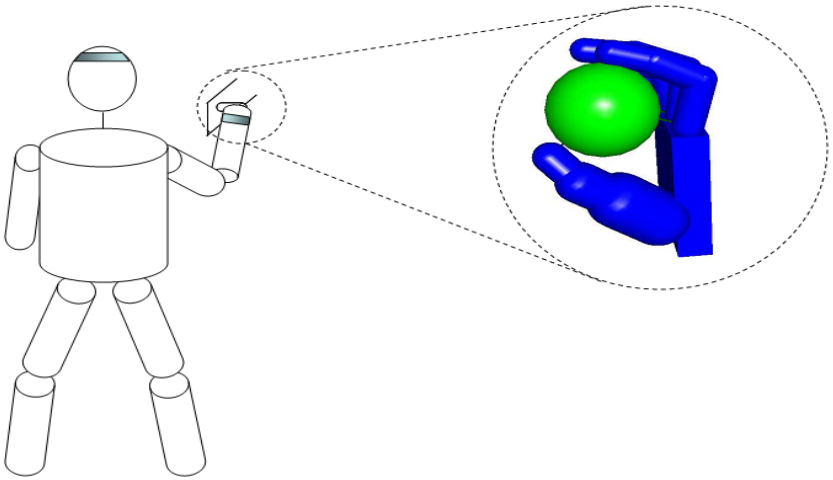


- Trained machine learning algorithm by millions of manually tagged images of people in different poses.
- OpenNI manages to adjust the most appropriate skeleton model to each human body in terms of size and pose.

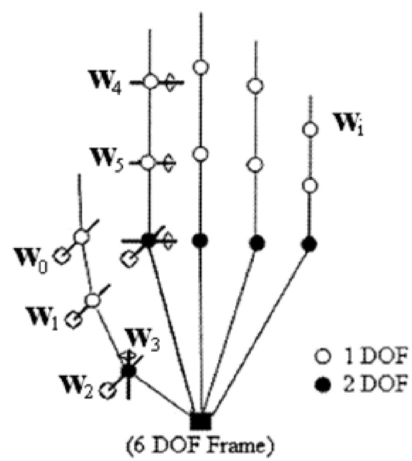
Imperial College
London

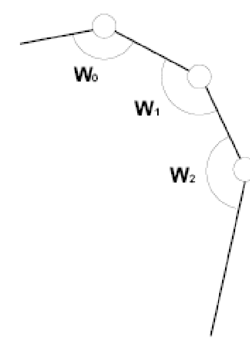
CeRTH
ITI

Cyberglove® finger tracker



- 4 DoF for each finger
- 3 phalanxes for each finger
- 3 DoF for the orientation of the whole palm
- Time-stamped Data





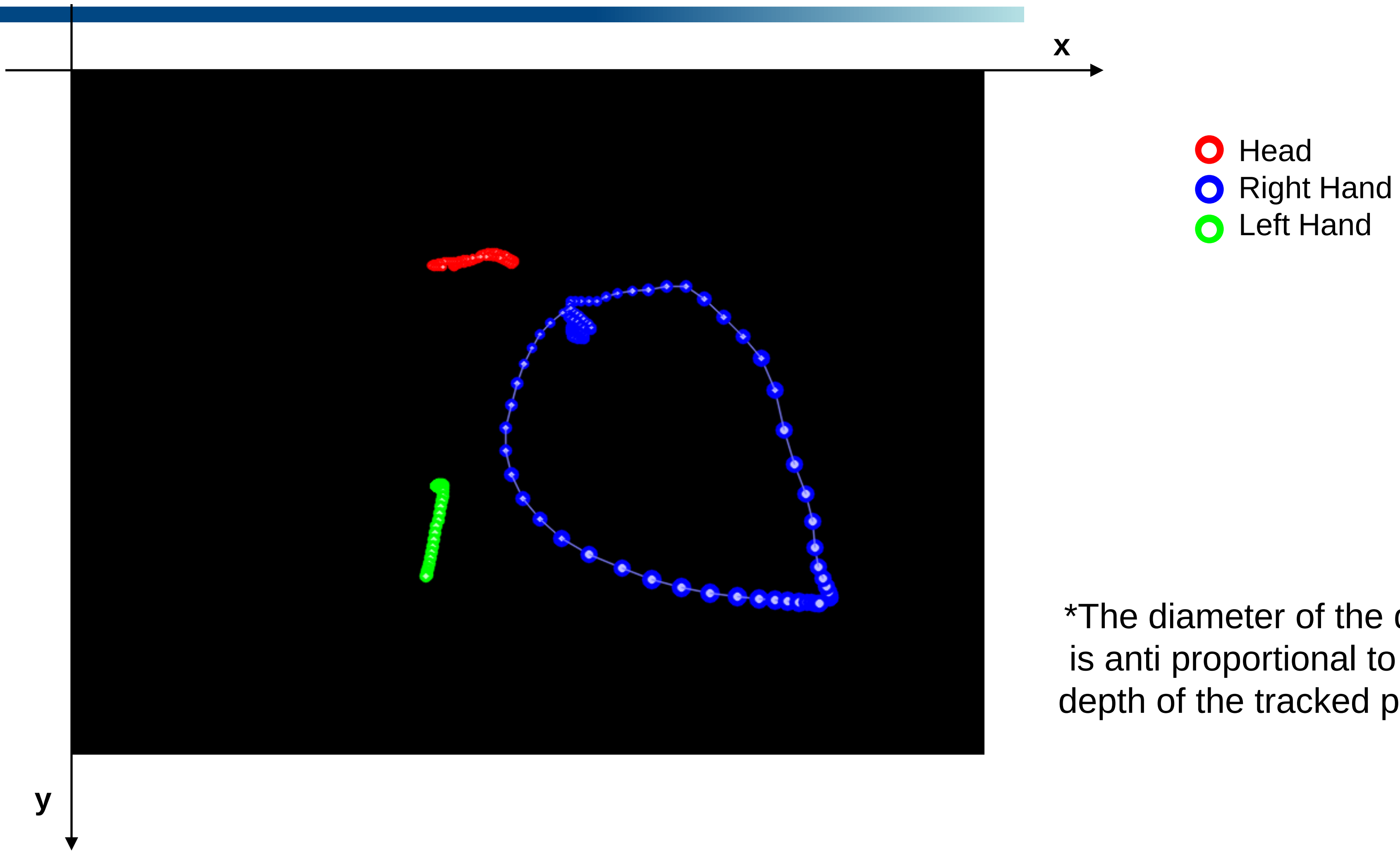
Outline

- Introduction
- Activity Detection
- Activity related recognition
  - Human Tracking
  - **Reaching Part**
  - Grasping Part
- Classifiers
- Results
- Conclusions

Imperial College London

CeRTH ITI

Feature Extraction (I)

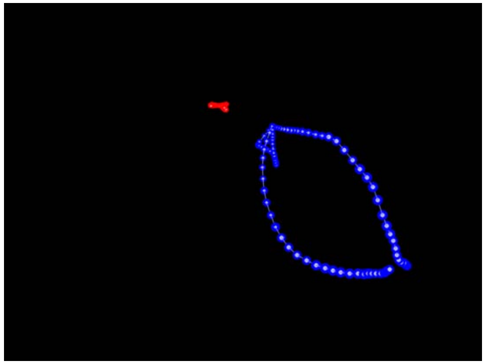
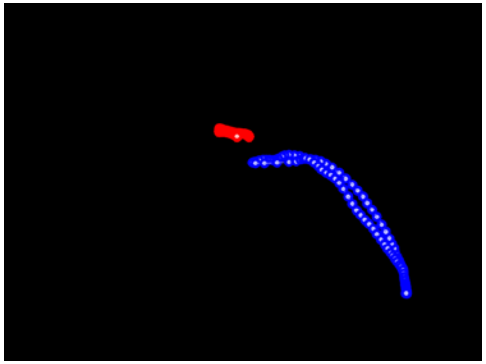
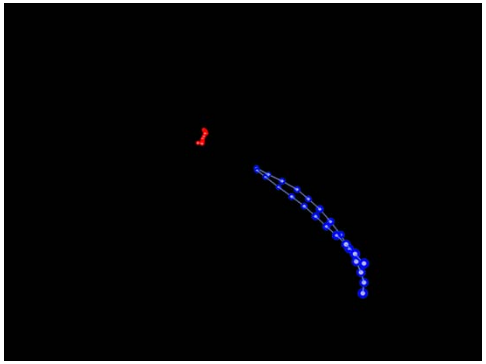
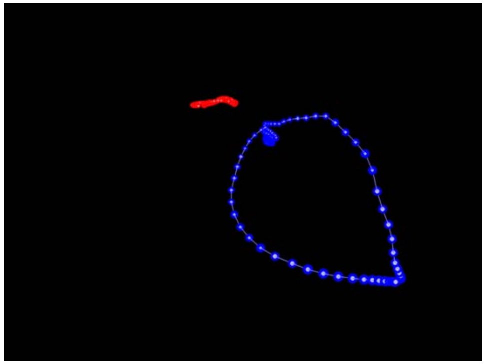
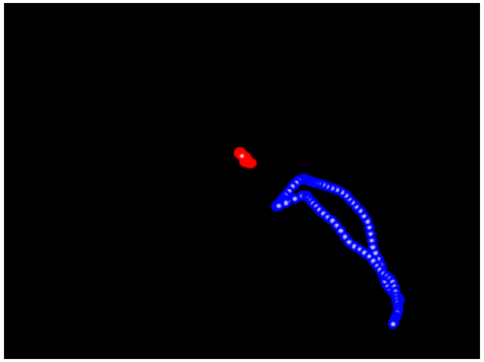
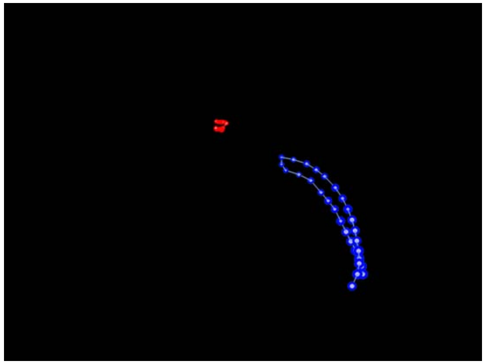


*The diameter of the dots is anti proportional to the depth of the tracked point.

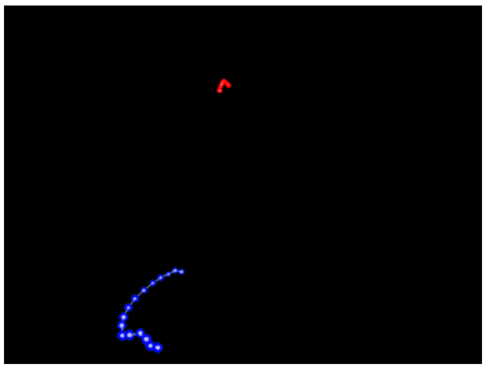
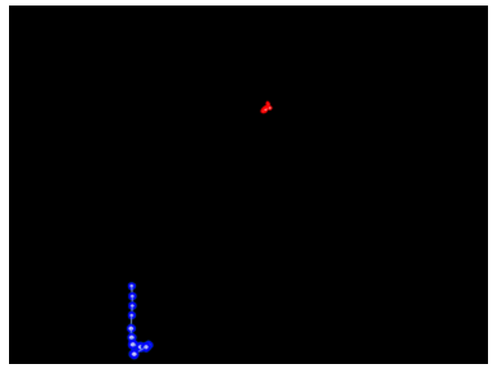
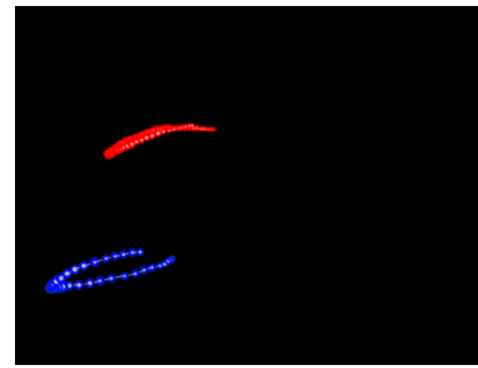
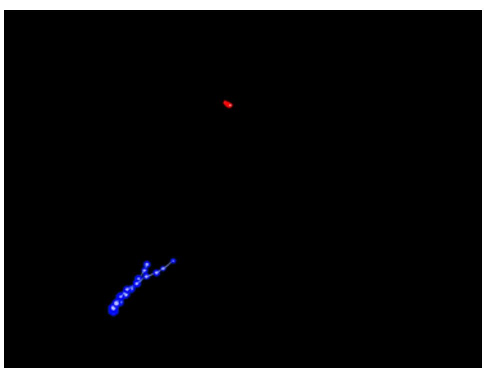
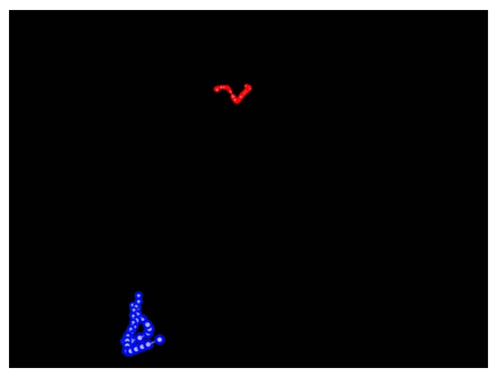
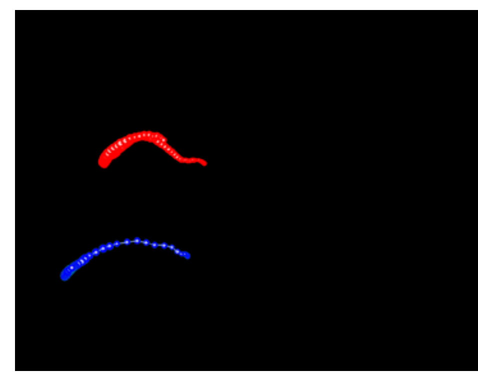
Imperial College London

CeRTH ITI

Phone Conversation Trajectories

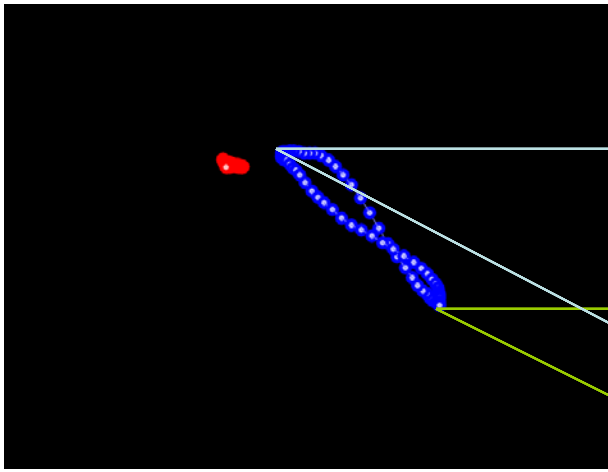
	Subject 1	Subject 2	Subject 3
1			
2			

Office Panel Trajectories

	Subject 1	Subject 2	Subject 3
1			
2			

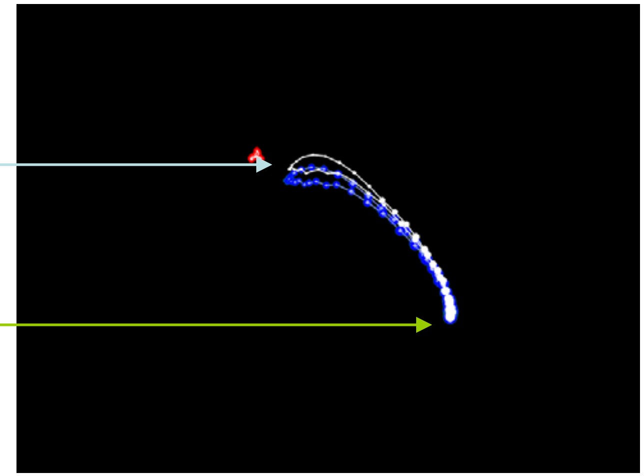
Warping the trajectory

Enrollment

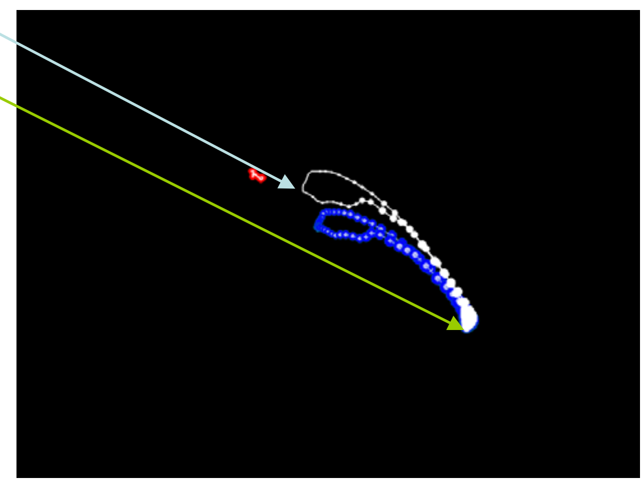


Authentication

Client:

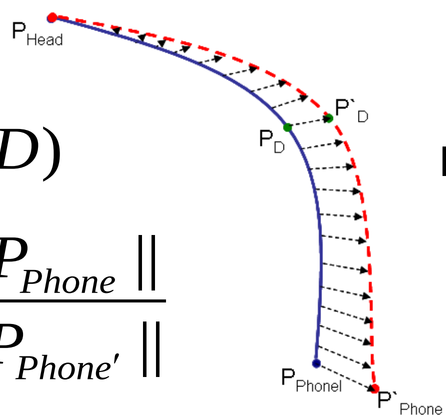


Impostor:



$$P(D') = q \cdot P(D)$$

$$q = \frac{\| P_{Head} - P_{Phone} \|}{\| P_{Head} - P_{Phone'} \|}$$



Feature Extraction (I)

- Exact location of the head/hand 3D position, at time *t (*<u>timestamp</u> tracking).

$$v_{x,y,z} = \frac{ds_{x,y,z}}{dt} \qquad a_{x,y,z} = \frac{dv_{x,y,z}}{dt}$$

- Dynamic Spatial Cost

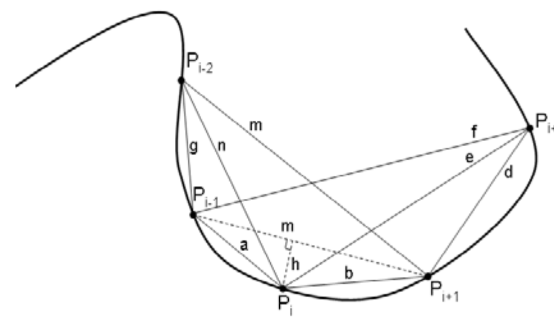
$$S_p(t) = S_p(t-1) + \sqrt{(x_t - x_{(t-1)})^2 + (y_t - y_{(t-1)})^2 + (z_t - z_{(t-1)})^2}$$

Imperial College London

CeRTH ITI

Feature Extraction (II)

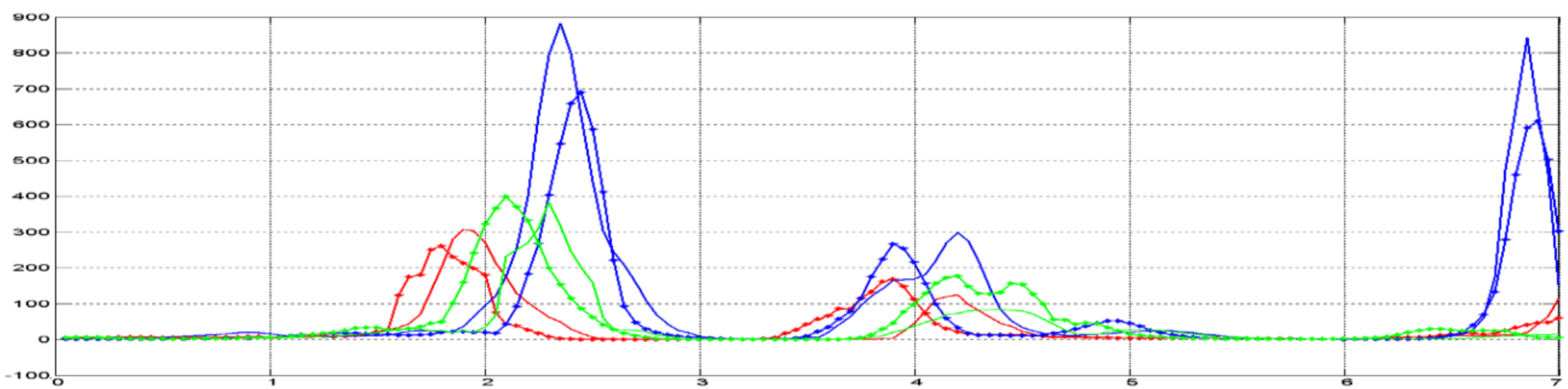
- Curvature - Derivative



$$\kappa^*(P_i) = 4\frac{\sqrt{\hat{s}(\hat{s}-a)(\hat{s}-b)(\hat{s}-c)}}{abc} \qquad \kappa_s^*(P_i) = 3\frac{\kappa^*(P_{i+1}) - \kappa^*(P_{i-1})}{2a+2b+d+g}$$

- Torsion - Derivative

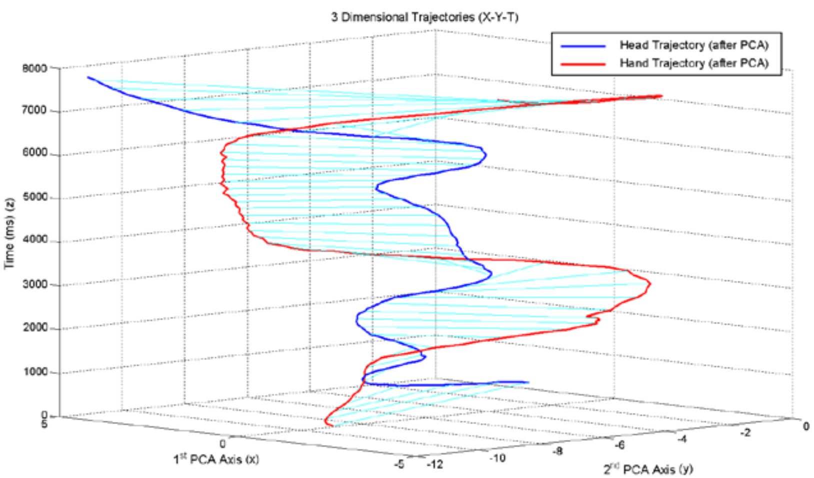
$$\tau^*(P_i) = \frac{1}{2}(6\frac{H^+}{def \cdot \kappa^*(P_i)} + 6\frac{H^-}{gmm \cdot \kappa^*(P_i)}) \qquad \tau_s^*(P_i) = 4\frac{\tau^*(P_{i+1}) - \tau^*(P_{i-1}) + r(\tau^*(P_i)\kappa_s^*(P_i)/6\kappa^*(P_i))}{2a+2b+2d+h+g}$$



Imperial College London

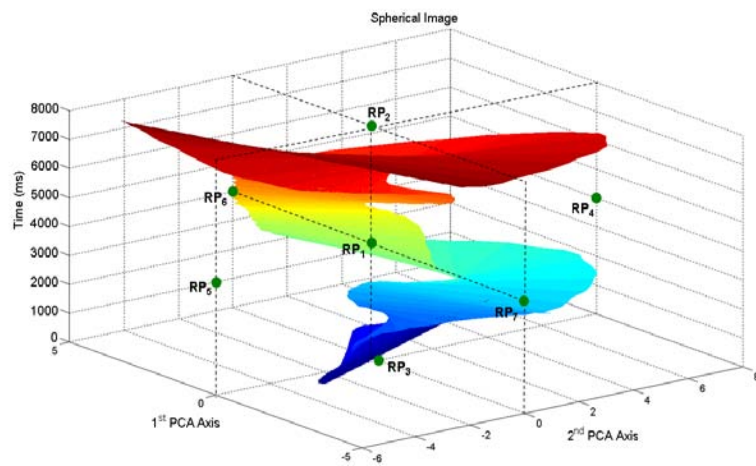
CeRTH ITI

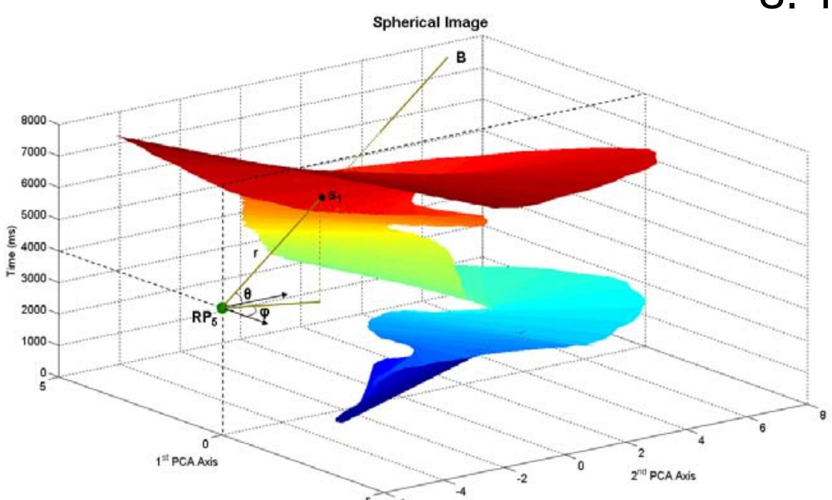
Spatiotemporal Activity Surface



1. PCA on 3D Trajectories
2. 3-dimensional manifold trajectories using time as 3rd dimension

3. Triangulation

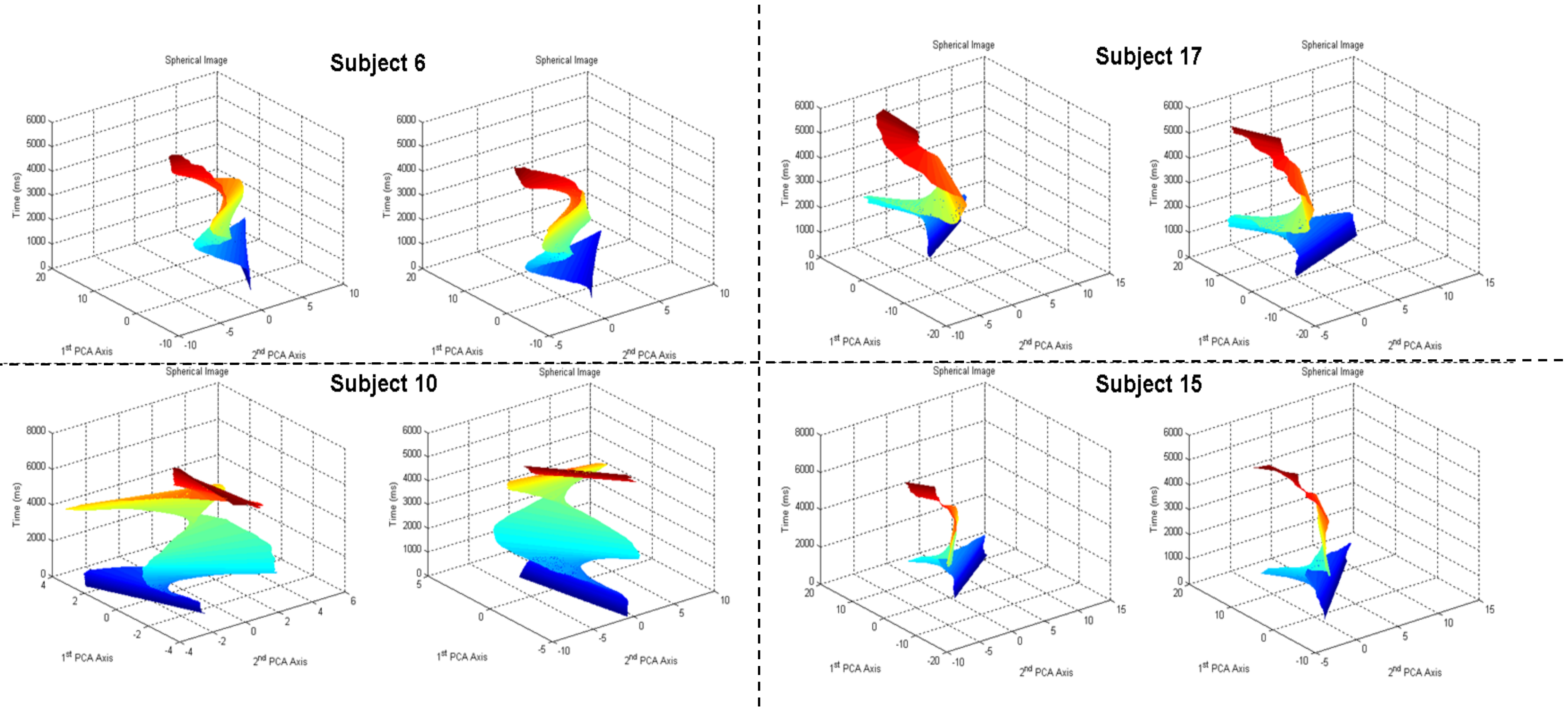




$$f_{RP_i}(\theta,\phi) = \min_{k=1,\ldots,K}\{d(RP_i, s_k)\}$$

Imperial College London
CeRTH ITI

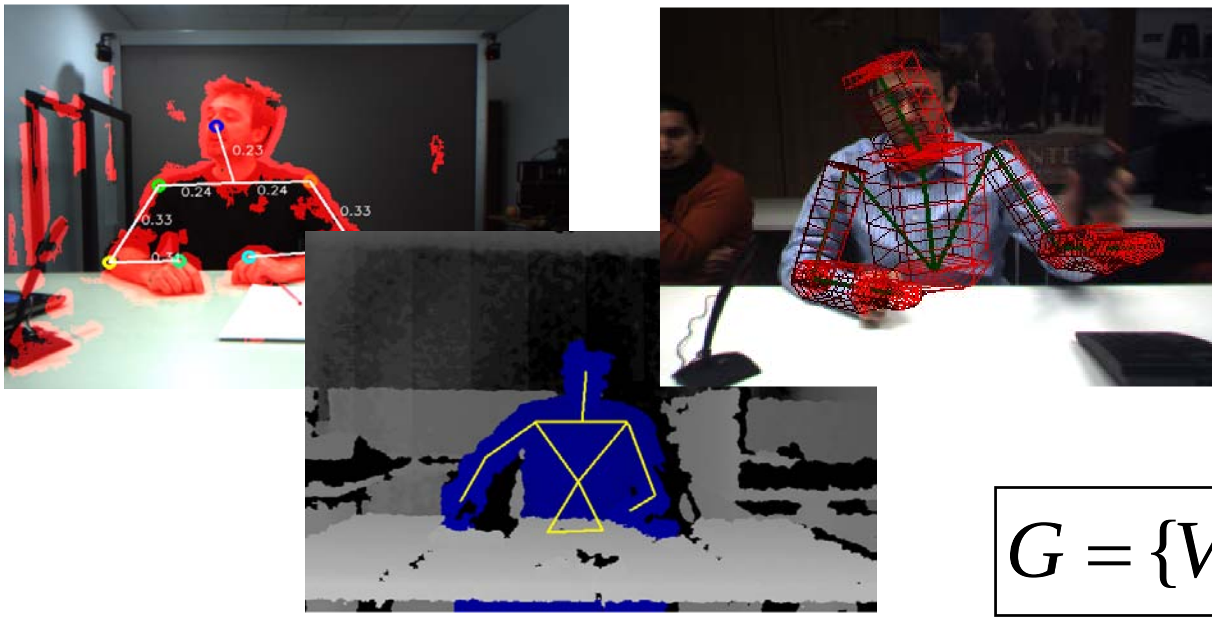
Intra-similarity vS. Inter-variance of AS



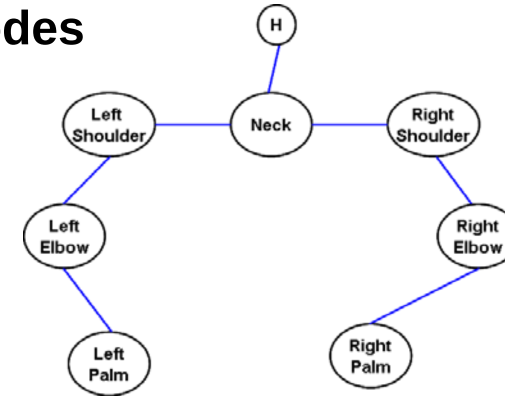
Imperial College London

CeRTH ITI

Static Anthropometric – Attributed Graph Matching



7 Attributed Edges
8 Nodes



$$\boxed{G = \{V, E, \{\mathrm{A}\}, \{\mathrm{B}\}_{i=1}^{7}\}}$$

Supports:
- Full-graph matching (n`=n)
- Sub-graph matching (n`>n)

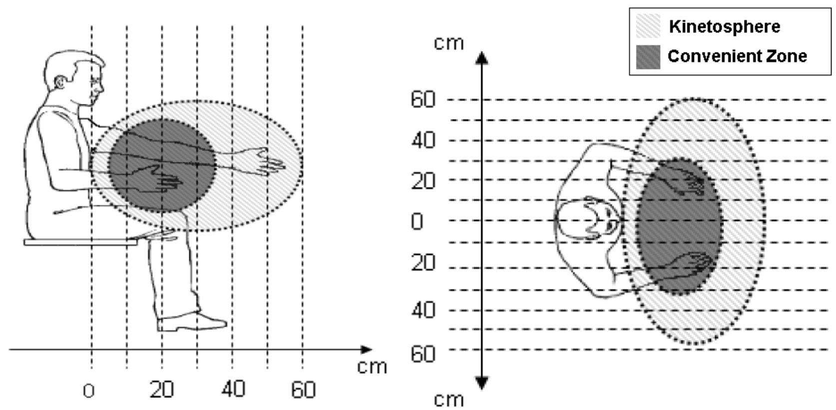
$$B_j = P_0 B' P_0^T{}_j (+\varepsilon M_j)$$

<u>(AGM) problem:</u>

$$\boxed{\min_p \left(\sum_{j=1}^{s} W_{j+r} \| B_j - PB'_j \|^q \right)}$$

Imperial College London
CeRTH ITI

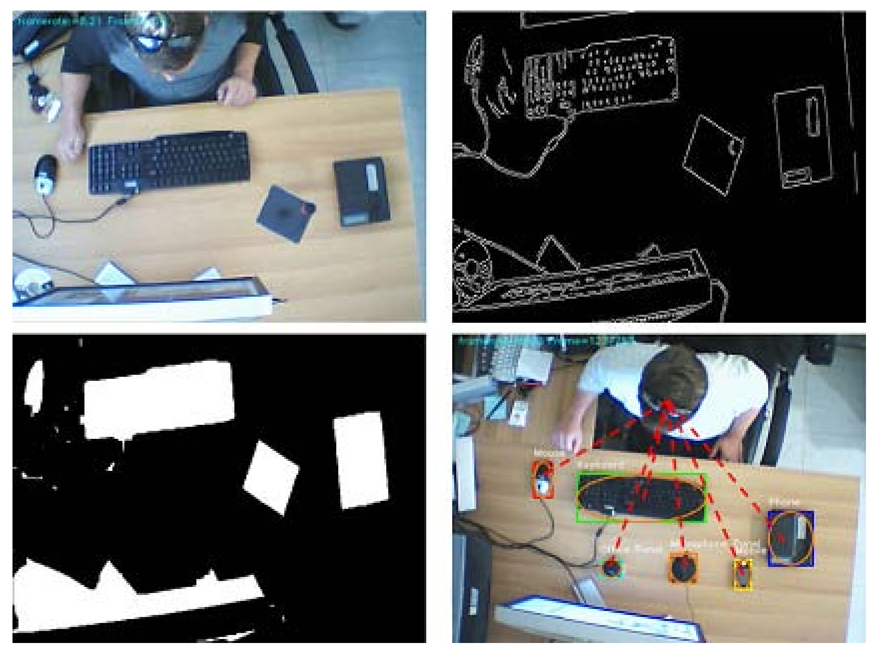
Confidence Factor – Ergonomy



$$b = \begin{cases} 0.1 \cdot d_{torso,object} + 0.5, \text{ if } d_{torso,object} < 5cm \\ 1, \text{ if } 5cm \le d_{torso,object} \le 35cm \\ -0.02 \cdot d_{torso,object} + 1.7, \text{ if } d_{torso,object} > 35cm \end{cases}$$

$$f_q = 1 - \frac{N_{missHead} + N_{missRHand} + N_{missLHand}}{3N_{frames}}$$

$$\boxed{f_{q,\,final} = b \cdot f_q}$$



Outline

- Introduction
- Activity Detection
- Activity related recognition
  - Human Tracking
  - Reaching Part
  - **Grasping Part**
- Classifiers
- Results
- Conclusions

Imperial College London

CeRTH

ITI

Feature Extraction

$$\overrightarrow{F_a} = \{\vec{\theta}_0, \vec{\theta}_1, \ldots, \vec{\theta}_{22}\}$$

- Angle between the phalanxes of each finger, at time *t* (timestamp tracking).

$$\omega_\theta = \frac{d\theta}{dt} \qquad a_\theta = \frac{d\omega}{dt} = \frac{d^2\theta}{dt^2}$$

- Dynamic Travel Cost

$$V_j(a_j(t), T_j) = \left(\frac{k_j a_j}{r}\right)\left(1 + \frac{[T_j - T_j^*(a_j)]^2}{s^2}\right)$$

The cost of moving joint j through an angle of size in the given time

$$T_j^*(a_j(t)) = k_j ln(a_j + 1),\ k_j \geq 0$$

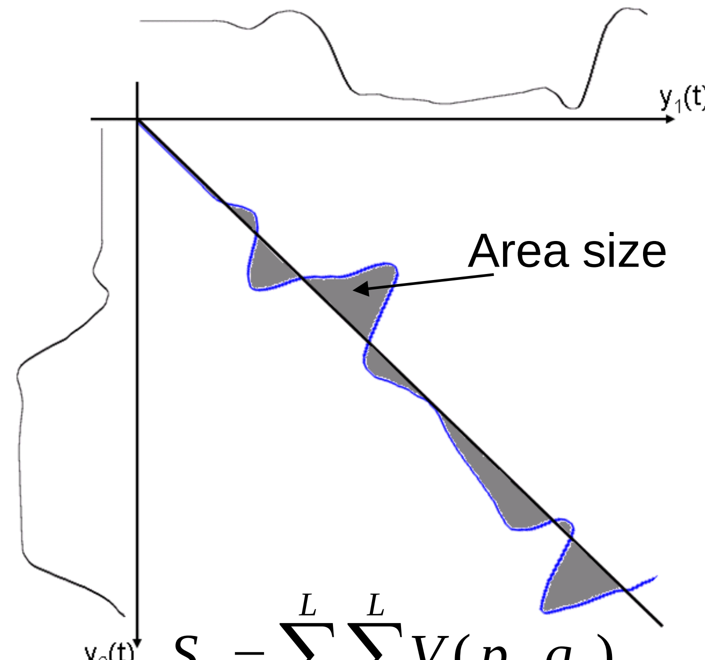
Joint's optimal time

Imperial College London

CeRTH ITI

Dynamic Time Warping (DTW) Classifier

- Based on dynamic Programming
- No training required
- Simple Implementation
- Fast



$$S_c = \sum_{i=1}^{L} \sum_{j=1}^{L} V(p_i, q_j)$$

• The problem of finding the optimal warping path can be reduced to finding this sequence of nodes (x_k; y_k), which minimizes D(x_k; y_k) along the complete path.

• The main aim is to find the path for which the least cost is associated.

$$D(x_k, y_k) = D(x_{(k-1)}, y_{(k-1)}) + c(x_k, y_k) = \sum_{m=1}^{k} c(x_m, y_m)$$

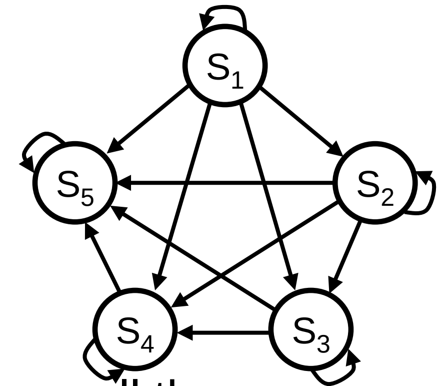
Total dissimilarity

$$D_M = S_c \cdot D_{min}(L, L)$$

Imperial College London

CeRTH ITI

Hidden Markov Model (HMM) Classifier

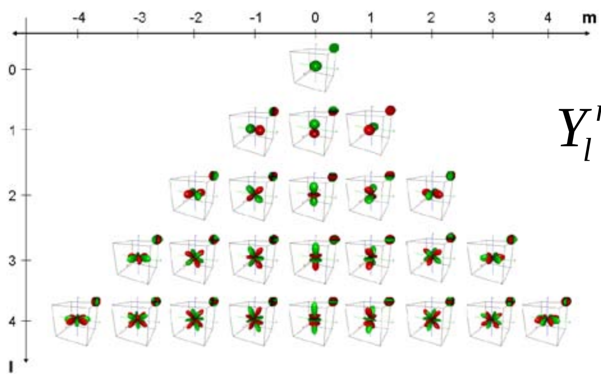


- Fully connected, left-to-right, five-state (N=5) HMM.
- The triplet: $$\lambda = \{\pi_j, \alpha_{ij}, b_j\}$$
  - π_j is the probability of the j^{th} state being the first state among all the trajectories,
  - a_{ij} is the probability of the j^{th} state occurring immediately after the i^{th} state,
  - b_j denotes the PDF of the j^{th} state.
- The observational data from each state of the HMM are generated according to a PDF dependent on the instant of t^{th} state, $b_j(O_t)$.
- Given HMMs for the *L* enrolled subjects and the new trajectory vectors, we assign user label *m* as the HMM that maximizes the likelihood (ML principle).

Imperial College London

CeRTH ITI

Spherical Harmonics



$$Y_l^m(\theta,\phi)=\sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^{|m|}cos(\theta)(cos(m\phi)+isin(m\phi))$$

Spherical Harmonics series

$$\boxed{c_m^l=\int_\Omega f_{RP_i}(\theta,\phi)Y_l^m(\theta,\phi)d\Omega=\int_0^{2\pi}\int_0^{\pi} f_{RP_i}(\theta,\phi)Y_l^m d\theta d\phi}$$

Spherical Harmonics coeffs.

Rotational Invariance: $c_l^*=\sum_{m=-K_l}^{K_l} c_l^m$

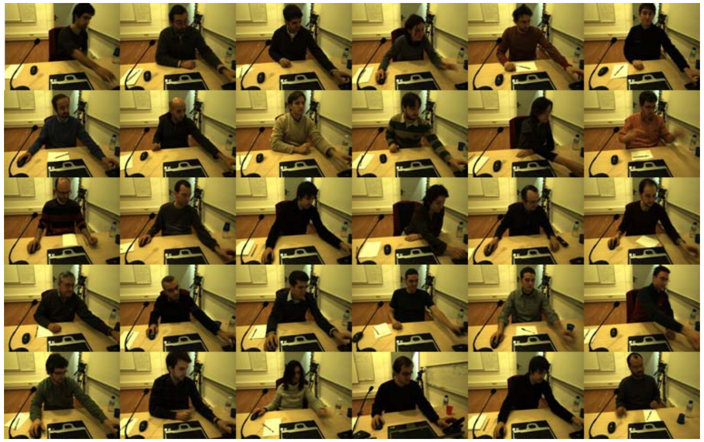
Z-Normalization: $c_l^z=\frac{c_l^*-\mu_y}{\sigma_y}$

Fusion among Reference Points: $S_{tot}=\sum_{j=1}^{N} w_j S_j = w_1 S_1 + w_2 S_2 + \ldots + w_N S_N$

Database

ACTIBIO dataset (AmI indoor environment):

- 29 Subjects,
- 2 Time sessions - 8 repetitions in total,
- Annotated frame sequences,
- 5 cameras in total:
  - 1 stereo camera (BumblebeeXB3 Point Grey Inc),
  - 2 usb cameras (Lateral-Zenithal)
  - …



Outline

- Introduction
- Activity Detection
- Activity related recognition
  - Human Tracking
  - Reaching Part
  - Grasping Part
- Classifiers
- **Results**
- Conclusions

Imperial College London

CeRTH ITI

Activity Detection Results

Events	*Phone*	*Panel*	*Mic. Panel*	*Drink*	*Hands Up*
Phone	**93.1%**	**0%**	**0%**	**6.9%**	**0%**
Panel	**0%**	**89.7%**	**10.3%**	**0%**	**0%**
Mic. Panel	**0%**	**3.44%**	**3.44%**	**93.1%**	**100%**
Drink	**0%**	**10.3%**	**86.2%**	**3.44%**	**0%**

Tracking Results

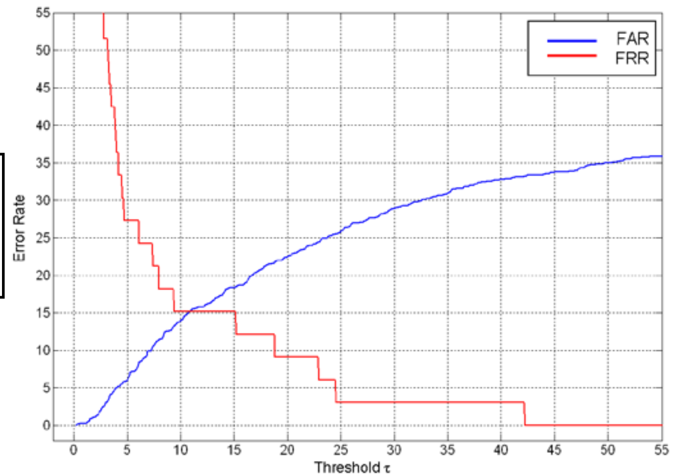
Subject’s No.	Confidence Score	Subject’s No.	Confidence Score
Subject 1	**0.835227**	*Subject 3*	**0.818681**
Subject 3	**0.937500**	*Subject 4*	**0.846774**
Subject 5	**0.933333**	*Subject 6*	**0.927778**
…	**…**	*…*	**…**
Subject 27	**0.935897**	*Subject 28*	**0.720588**
Subject 29	**0.866667**		

Imperial College London

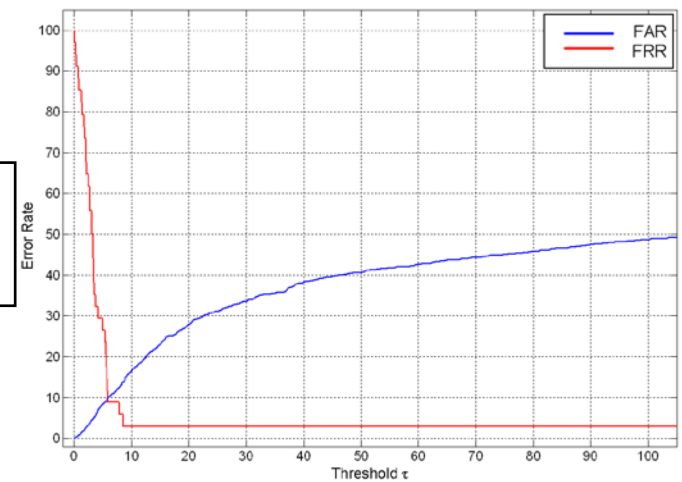
CeRTH ITI

Authentication HMM Results

Phone Conv. (**EER=15%**)

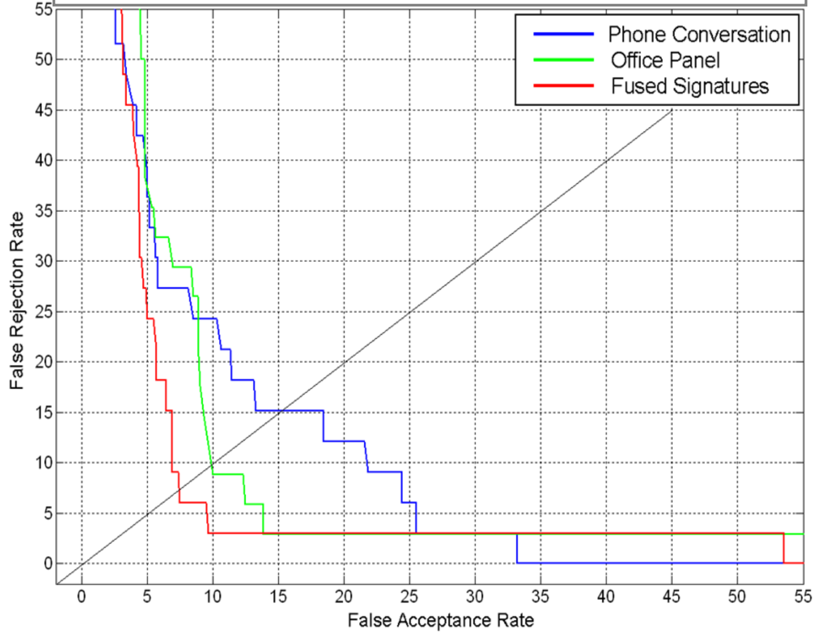


Office Panel (**EER=9.8%**)

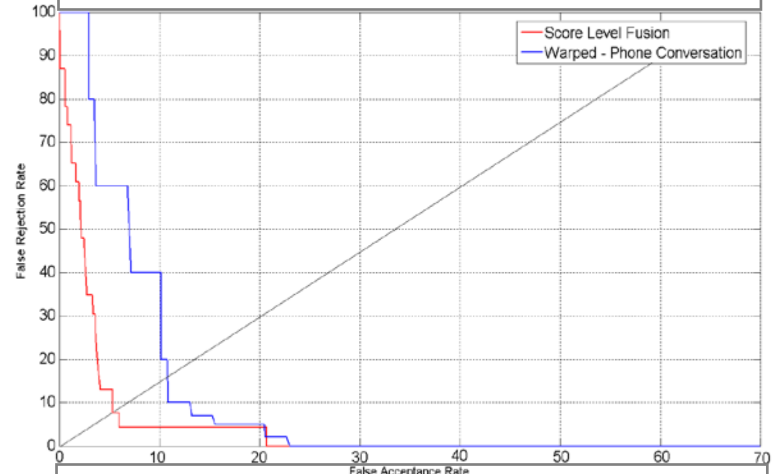


Authentication Performance

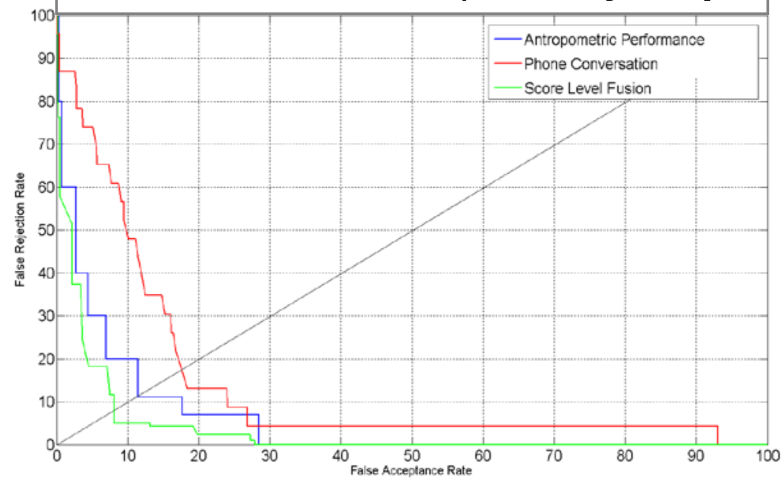
ACTIBIO Database (**29 Subjects)**



ITI Internal Database (**15 Subjects**



ACTIBIO Database (**29 Subjects)**



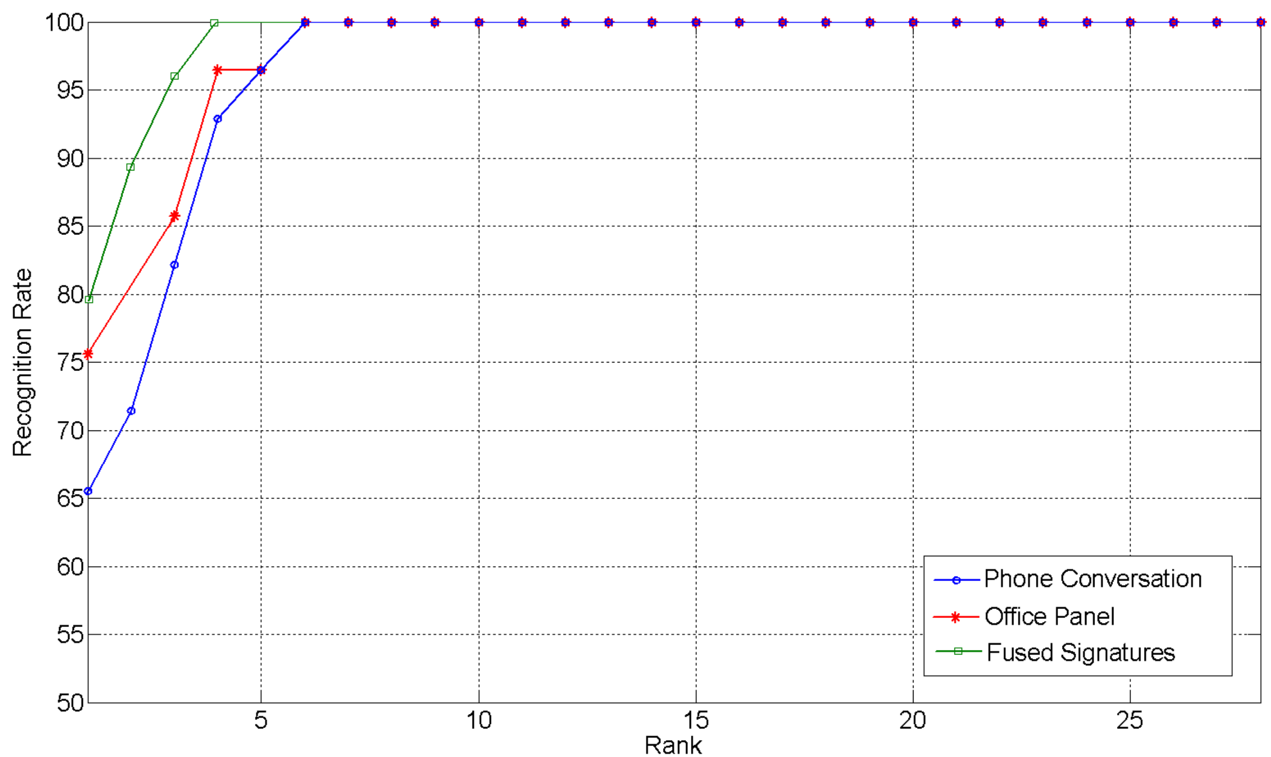
Score Level Fusion:

$$S_{total} = 0.25 * S_{Phone} + 0.75 * S_{Panel}$$

Imperial College London

CeRTH ITI

Recognition Performance (HMM)



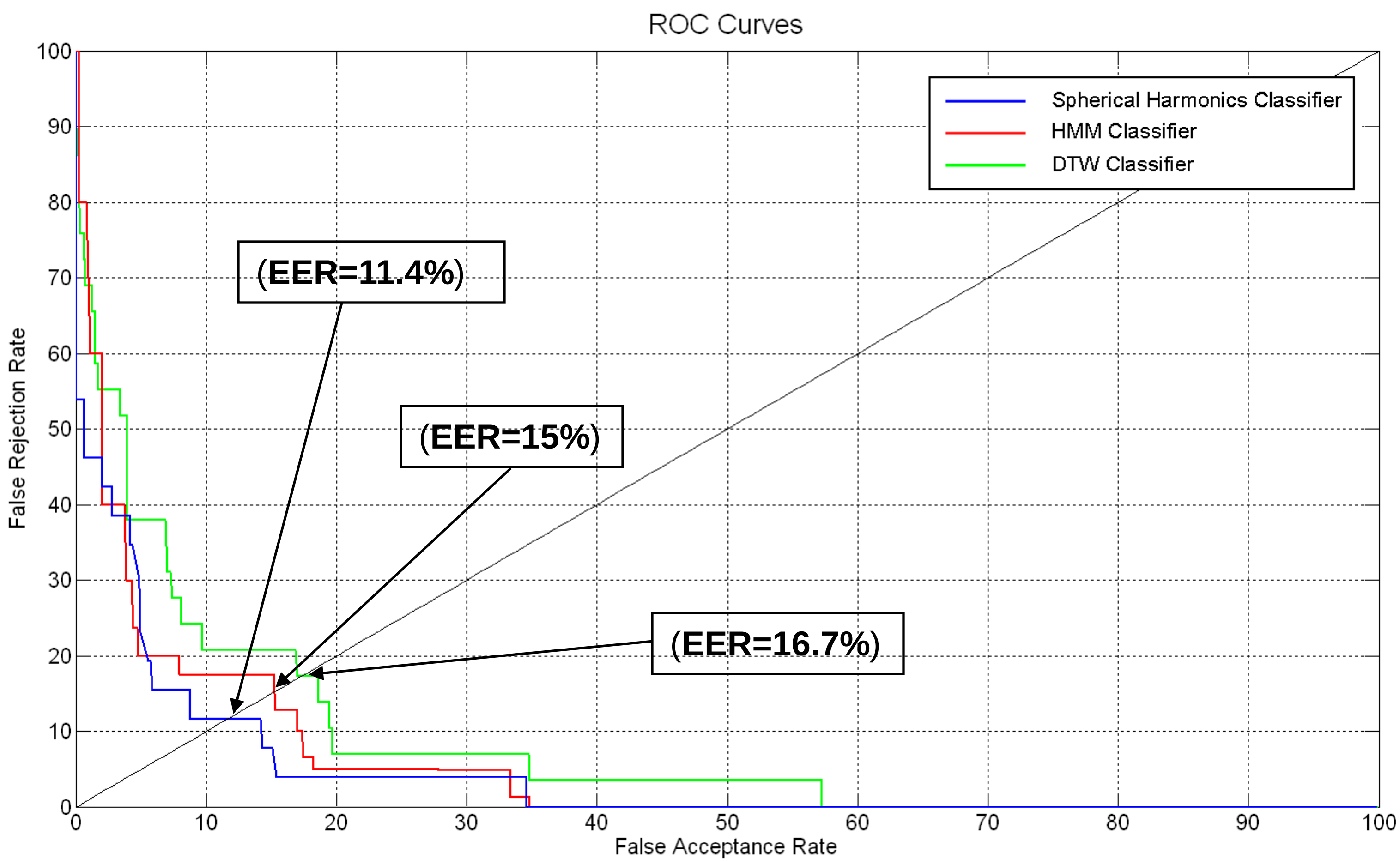
Score Level Fusion:

$$S_{total} = 0.25 * S_{Phone} + 0.75 * S_{Panel}$$

Imperial College London

CeRTH ITI

Spherical Harmonics Results



Conclusions

- Recognition using just a common stereoscopic camera.
- Unobtrusive, on the move, continuous authentication.
- Significant recognition potential just by the spatial information of the trajectories.
- Higher authentication rates are expected, given the relative entropies.
- Rotational invariance from spherical harmonics analysis.
- Privacy enabled method, since no obtrusive information is stored.
- Activity-related biometric authentication provided very promising results and is expected to maximize the performance of a multimodal biometric system.