Body and gestures: models, algorithms , applications

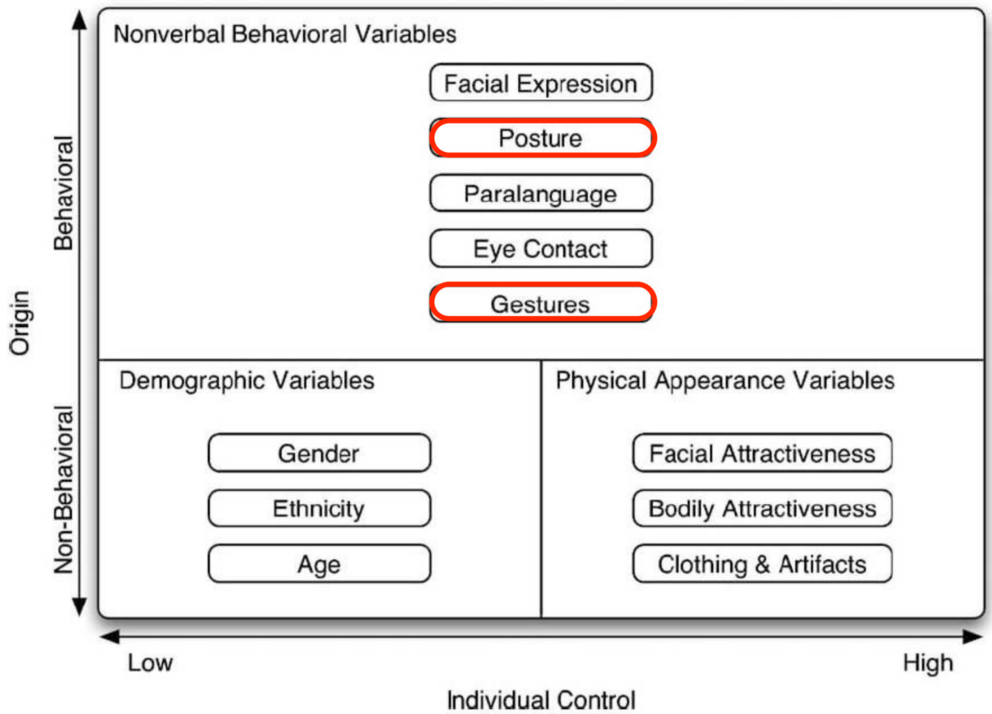
Corso di Interazione uomo-macchina II

Prof. Giuseppe Boccignone

Dipartimento di Informatica
Università di Milano

boccignone@di.unimi.it
http://boccignone.di.unimi.it/IUM2_2014.html

Non verbal behavior //De Meuse



DeMeuse (1987)

Posture and gesture interaction

	Example Social Behaviours							Tech.		
Social Cues	emotion	personality	status	dominance	persuasion	regulation	rapport	speech analysis	computer vision	biometry
Physical appearance										
height			√	√					√	√
attractiveness		√	√	√	√		√		√	√
body shape		√		√					√	√
Gesture and posture										
hand gestures	√	√			√	√	√		√	√
posture	√	√	√	√	√	√	√		√	√
walking		√	√	√					√	√
Face and eyes behaviour										
facial expressions	√	√	√	√	√	√	√		√	√
gaze behaviour	√	√	√	√	√	√	√		√	
focus of attention	√	√	√	√	√	√	√		√	
Vocal behaviour										
prosody	√	√		√	√		√	√		
turn taking	√	√	√	√		√	√	√		
vocal outbursts	√	√		√	√	√	√	√		
silence	√		√				√	√		
Space and Environment										
distance	√	√	√		√		√		√	
seating arrangement				√	√		√		√	

A. Vinciarelli, M. Pantic, H. Bourlard, *Social Signal Processing: Survey of an Emerging Domain*, Image and Vision Computing (2008)

Analysing postures and gestures

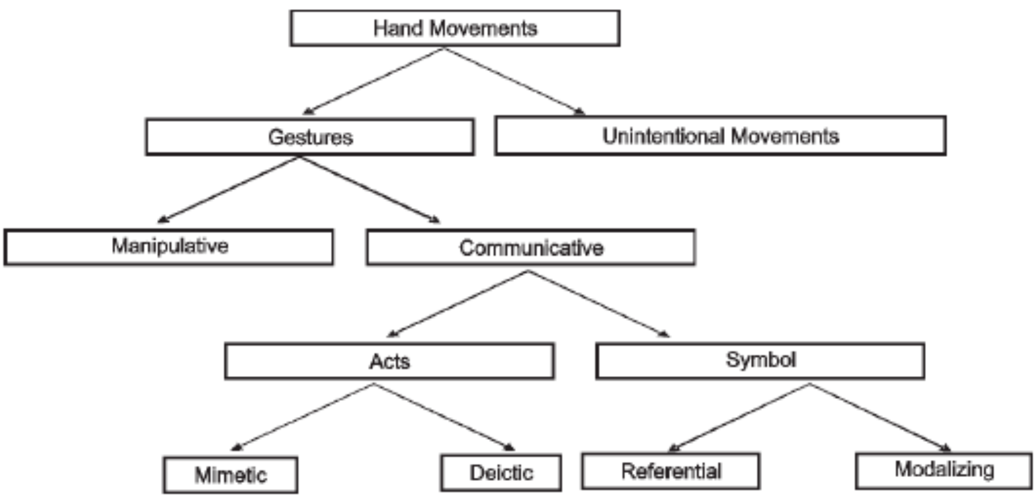
- The primary goal of gesture recognition research is to create a system which can identify specific human gestures and use them to convey information or for device control.
- A gesture may be defined as a physical movement of the hands, arms, face, and body with the intent to convey information or meaning.
- Gesture recognition, then, consists not only of the tracking of human movement, but also the interpretation of that movement as semantically meaningful commands

Analysing postures and gestures //application areas

General domain	Specific area
Virtual reality	—Interactive virtual worlds —Games —Virtual studios —Character animation —Teleconferencing (e.g., film, advertising, home-use)
"Smart" surveillance systems	—Access control —Parking lots —Supermarkets, department stores —Vending machines, ATMs —Traffic
Advanced user interfaces	—Social interfaces —Sign-language translation —Gesture driven control —Signaling in high-noise environments (airports, factories)
Motion analysis	—Content-based indexing of sports video footage —Personalized training in golf, tennis, etc. —Choreography of dance and ballet —Clinical studies of orthopedic patients
Model-based coding	—Very low bit-rate video compression

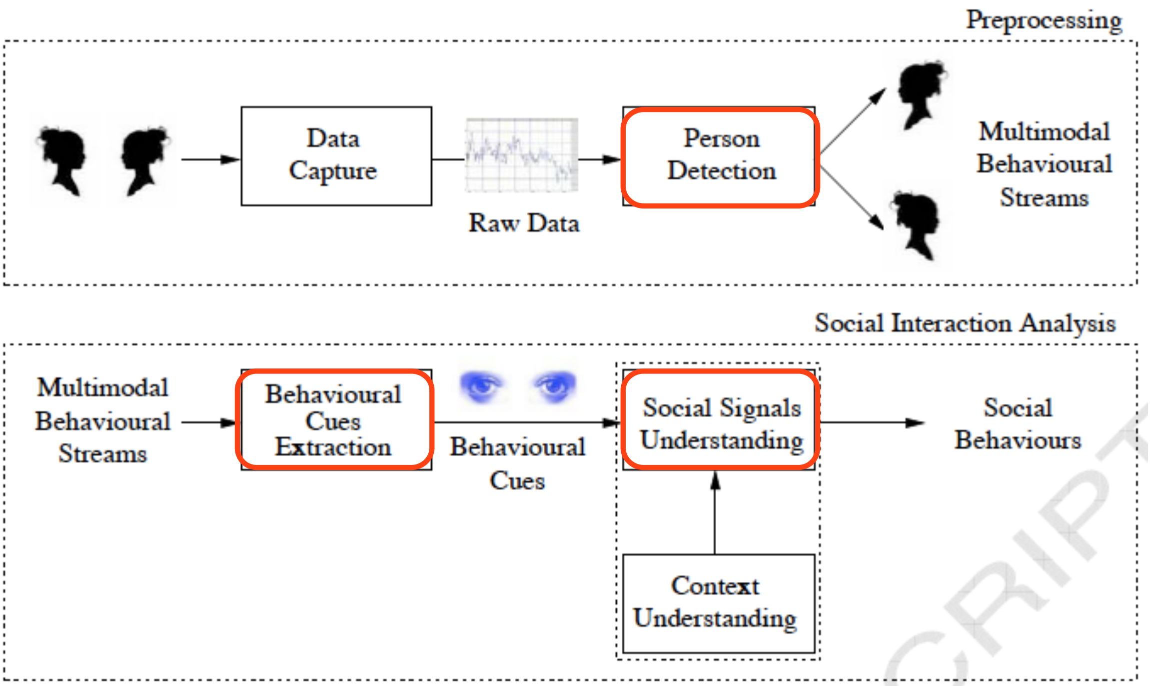
Hand gestures

- Taxonomy of hand gestures for HCI



- Visual interpretation of hand gestures can help in achieving the ease and naturalness desired for Human Computer Interaction (HCI).

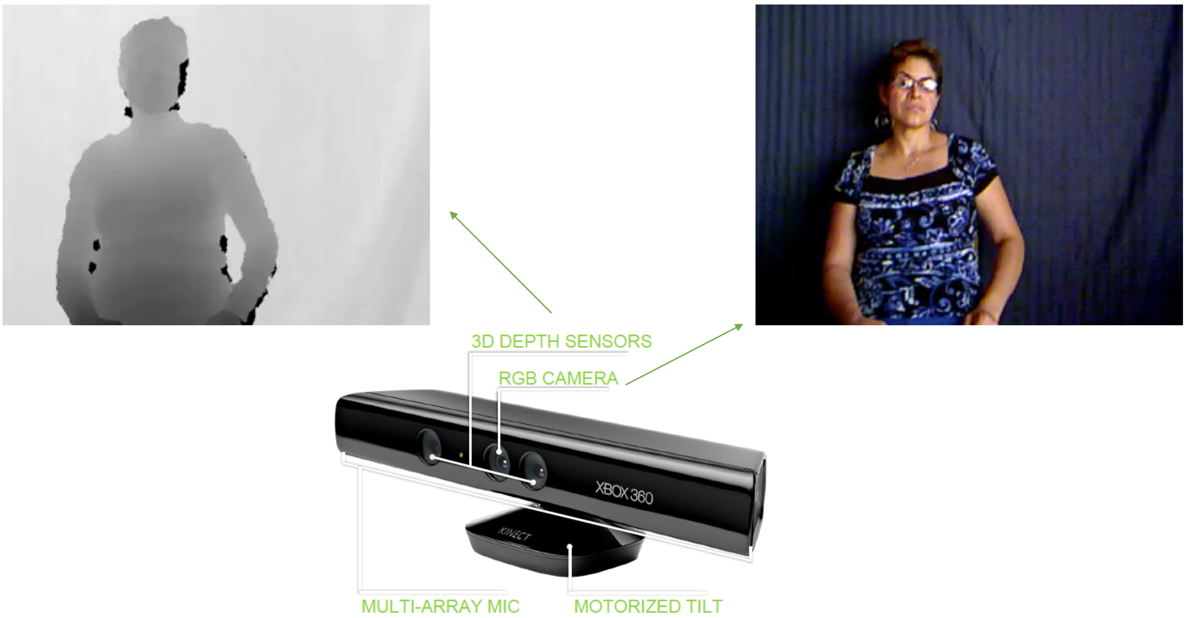
Posture and gesture interaction



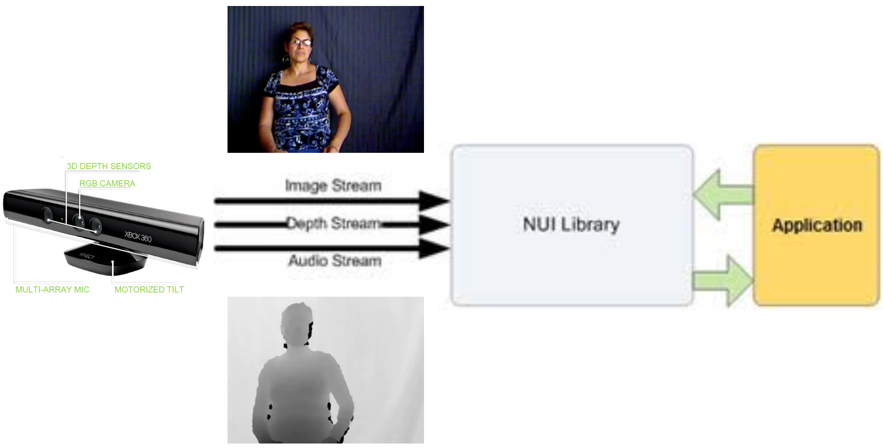
Analysing postures and gestures

- There are two main challenges in recognizing posture and gestures:
  - detecting the body parts involved in the gesture (e.g. hands)
  - modeling the temporal dynamics of the gesture

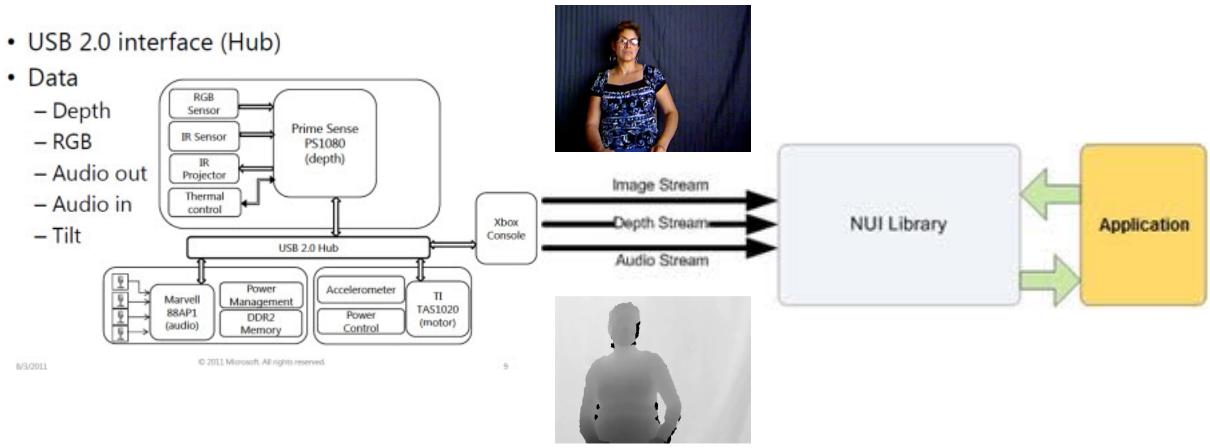
Analysing postures and gestures: //detecting parts: Kinect style



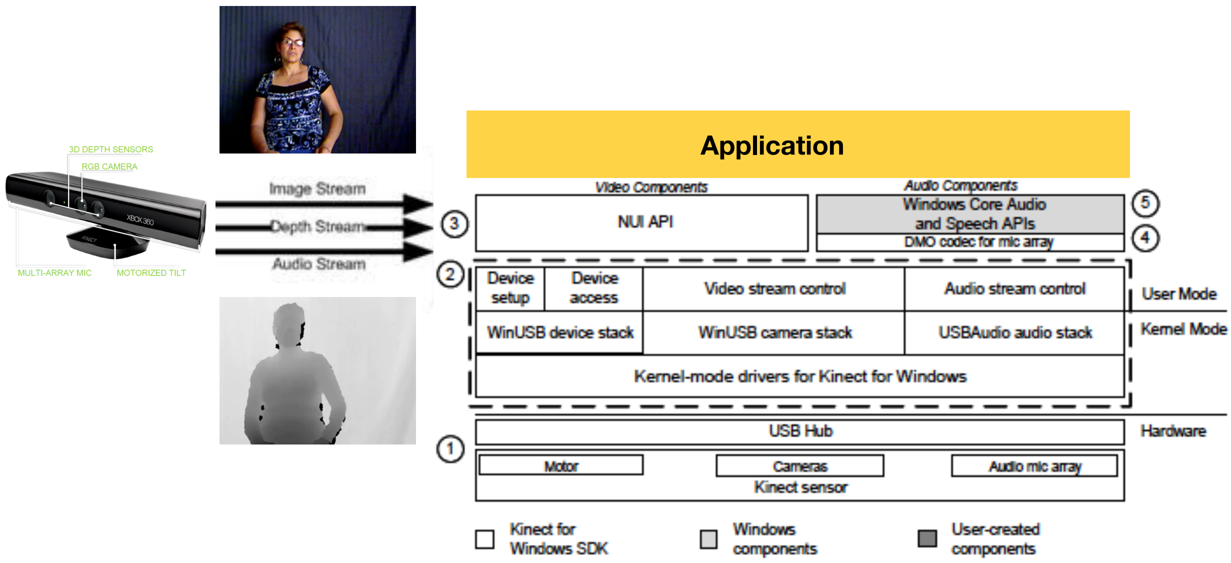
Analysing postures and gestures: //detecting parts: Kinect style (Microsoft)



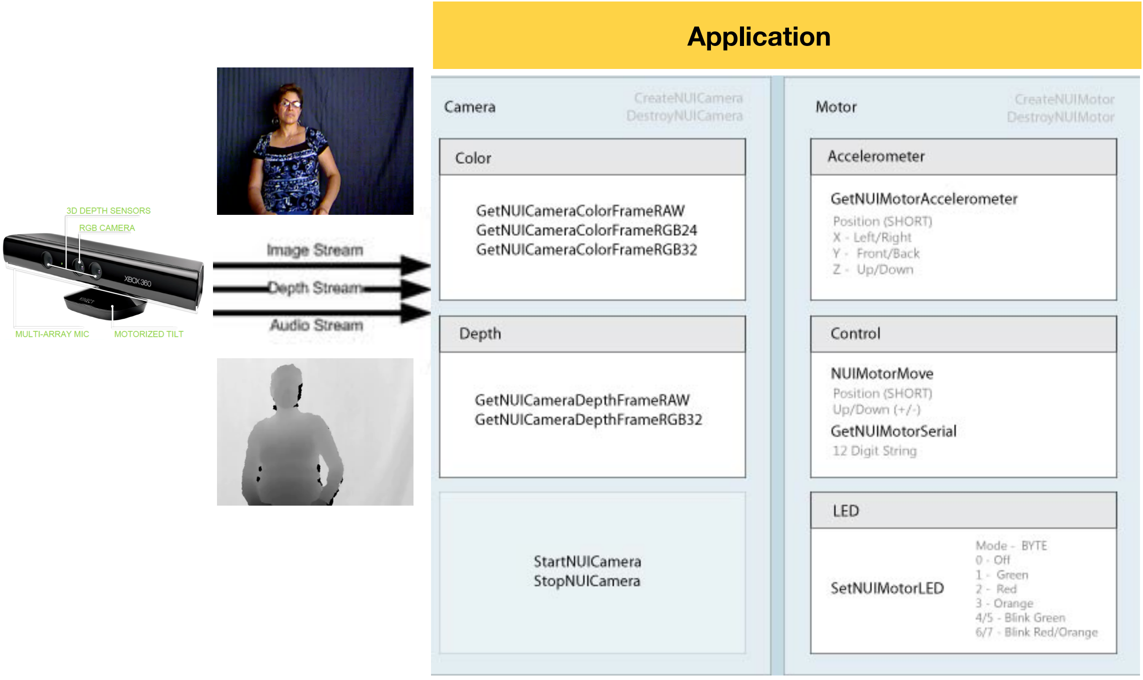
Analysing postures and gestures: //detecting parts: Kinect style (Microsoft)



Analysing postures and gestures: //detecting parts: Kinect style (Microsoft)



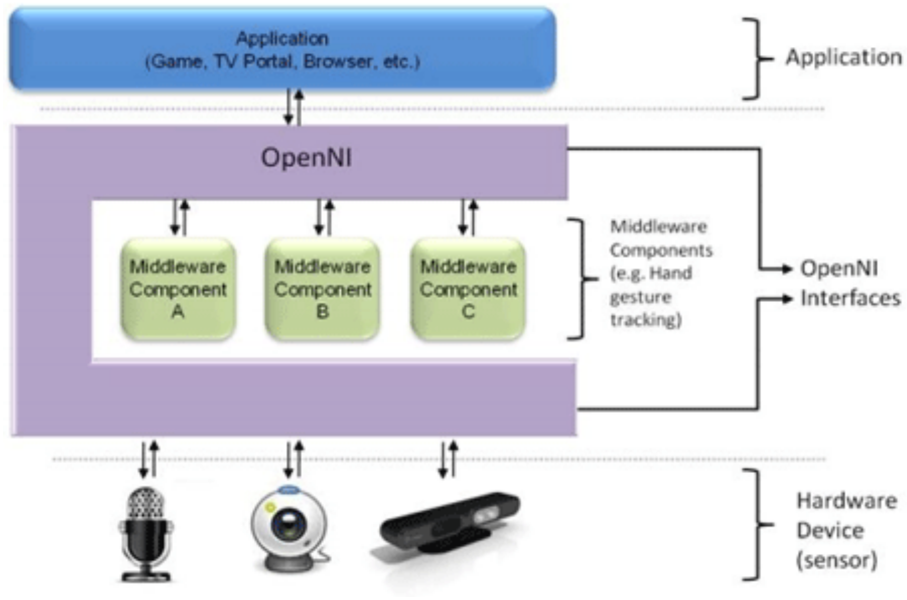
Analysing postures and gestures:
//detecting parts: Kinect style (Microsoft)



Analysing postures and gestures:
//detecting parts: Kinect style (OpenNI)

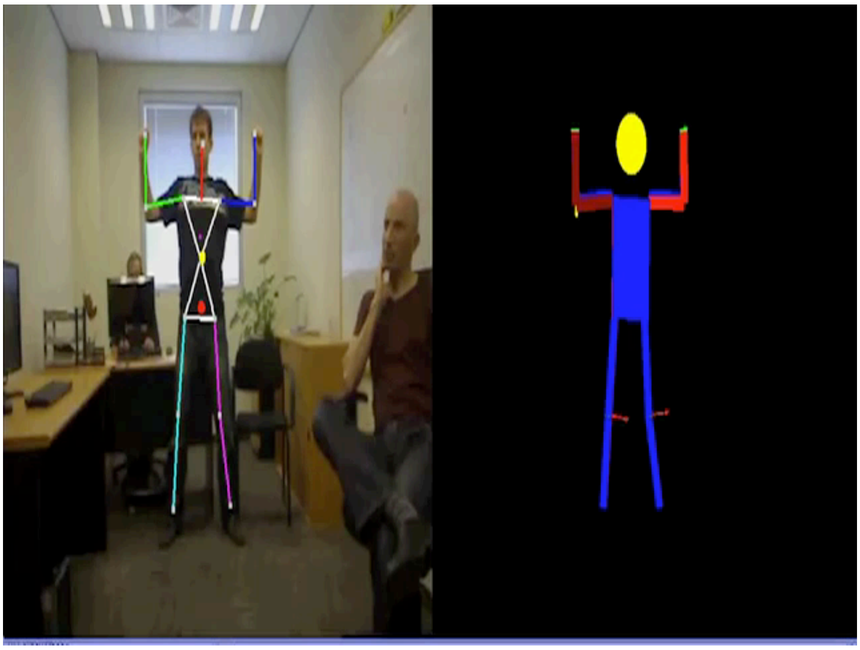
http://www.openni.org/

Kinect style:
//http://www.openni.org/

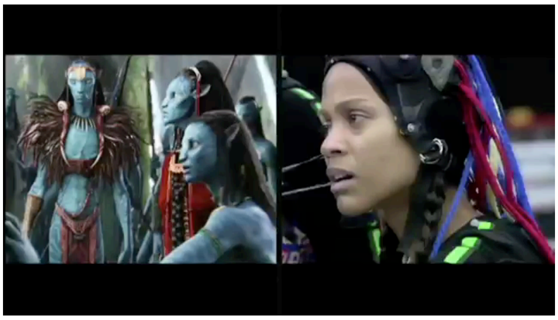


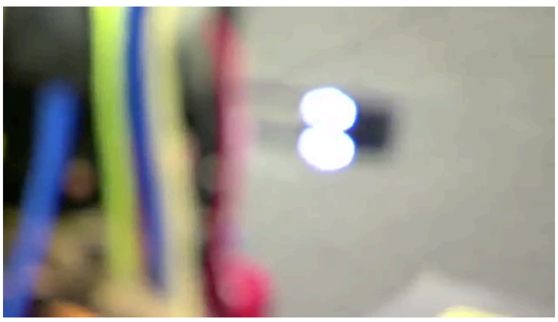
Kinect style:
//http://www.openni.org/

Kinect style:
//http://www.openni.org/



Analysing postures and gestures:
//detecting parts: motion capture





Analysing postures and gestures: //detecting parts: computer vision

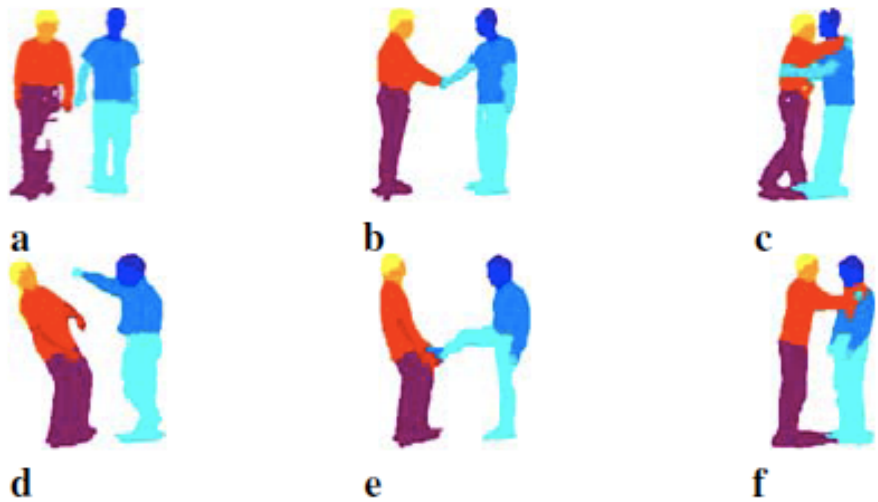
- There are two main challenges in recognizing posture and gestures:
  - detecting the body parts involved in the gesture (e.g. hands)
    - addressed by selecting appropriate visual features: these include, e.g., histograms of oriented gradients , optical flow, spatio-temporal salient points and space-time volumes .
  - modeling the temporal dynamic of the gesture

Analysing postures and gestures: //detecting parts: computer vision

Analysing postures and gestures:
//detecting parts: computer vision

tracked body parts indexed by different colors

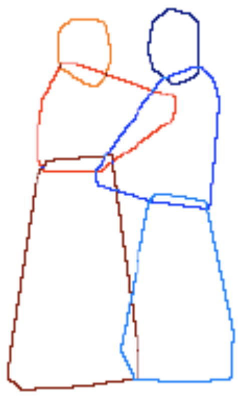
Analysing postures and gestures:
//detecting parts: computer vision



tracked body parts indexed by different colors

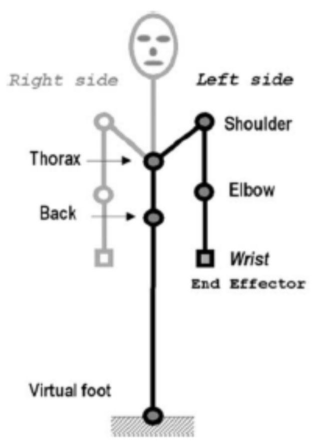
Analysing postures and gestures:
//detecting parts: computer vision

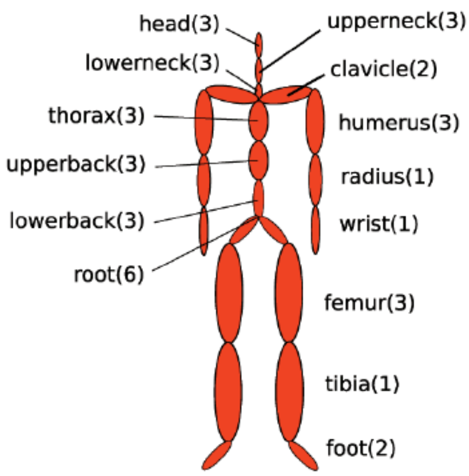
ellipse



convex hull

Analysing postures and gestures:
//representing parts: computer vision





Analysing hand gestures:
//representing parts: computer vision

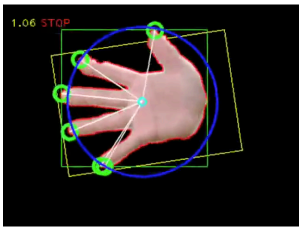
Spatial Gesture Model

3D Hand Model-Based

Parameters:
- joint angles
- palm position

Appearance-Based

Parameters:
- images
- image geometry parameters
- image motion parameters
- fingertip position & motion

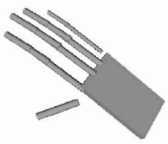


Analysing hand gestures:
//representing parts: computer vision

Contour

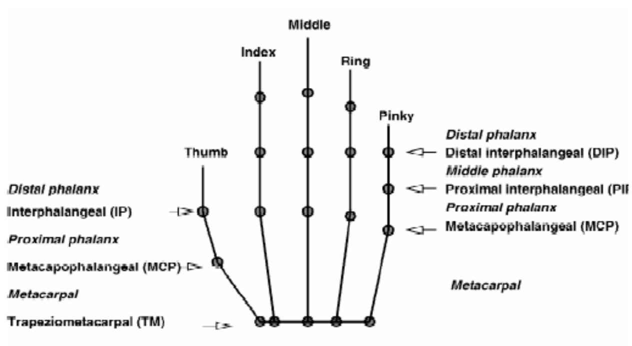
Binary silhouette.

3D skeletal model

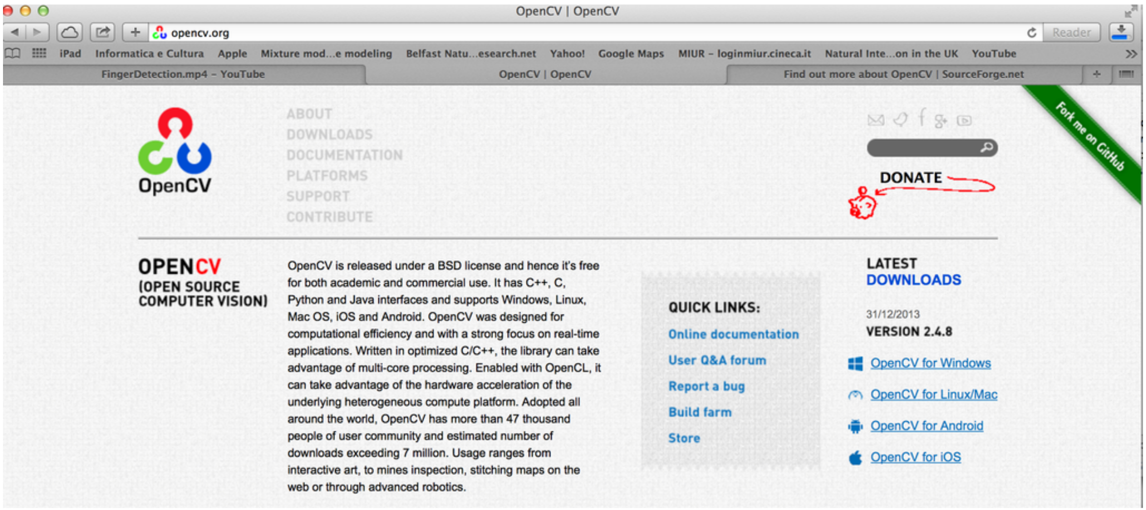


3D wireframe volumetric model.

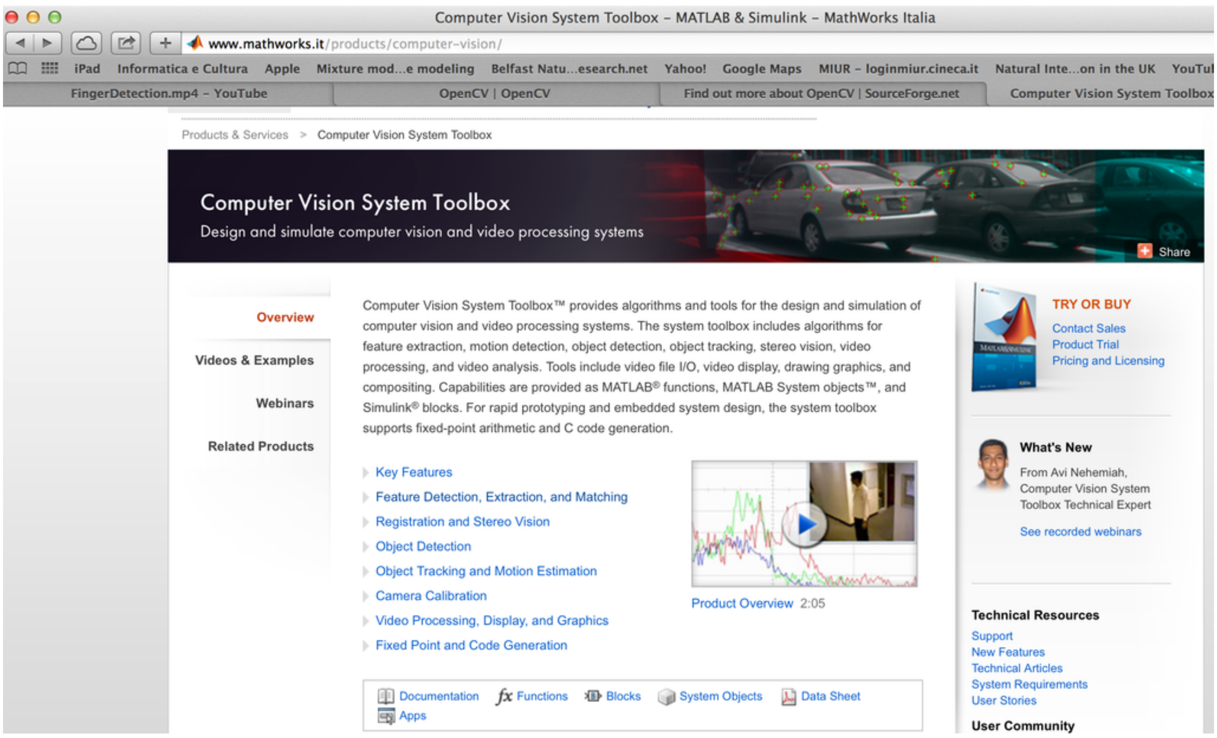
3D Textured volumetric model



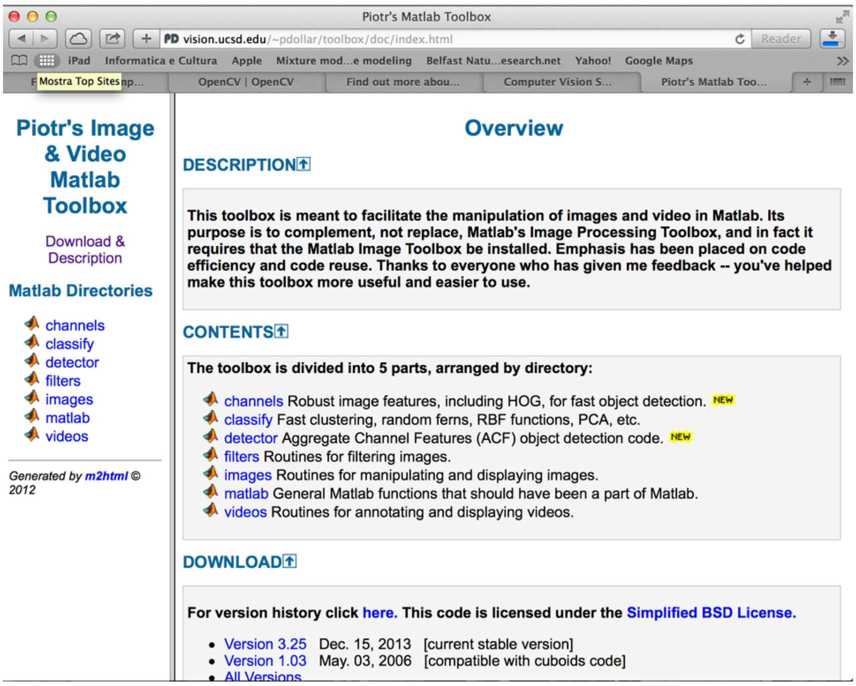
Analysing postures and gestures:
//detecting parts: computer vision (opencv)



Analysing postures and gestures:
//detecting parts: computer vision (Matlab)

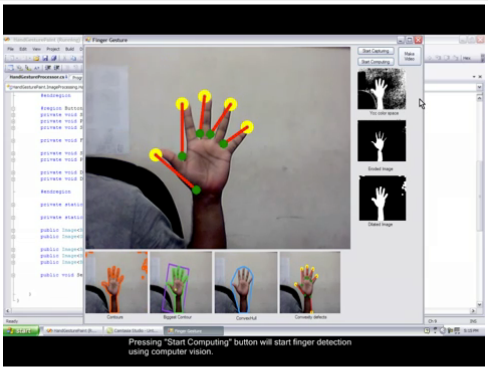


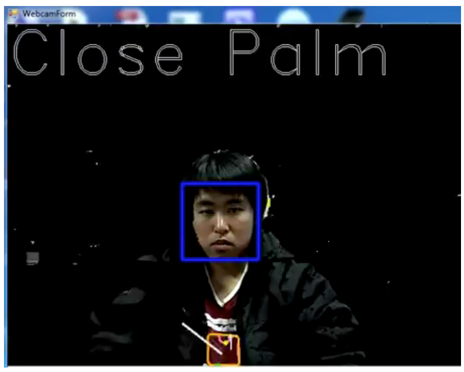
Analysing postures and gestures:
//detecting parts: computer vision (Piotr’s toolbox)



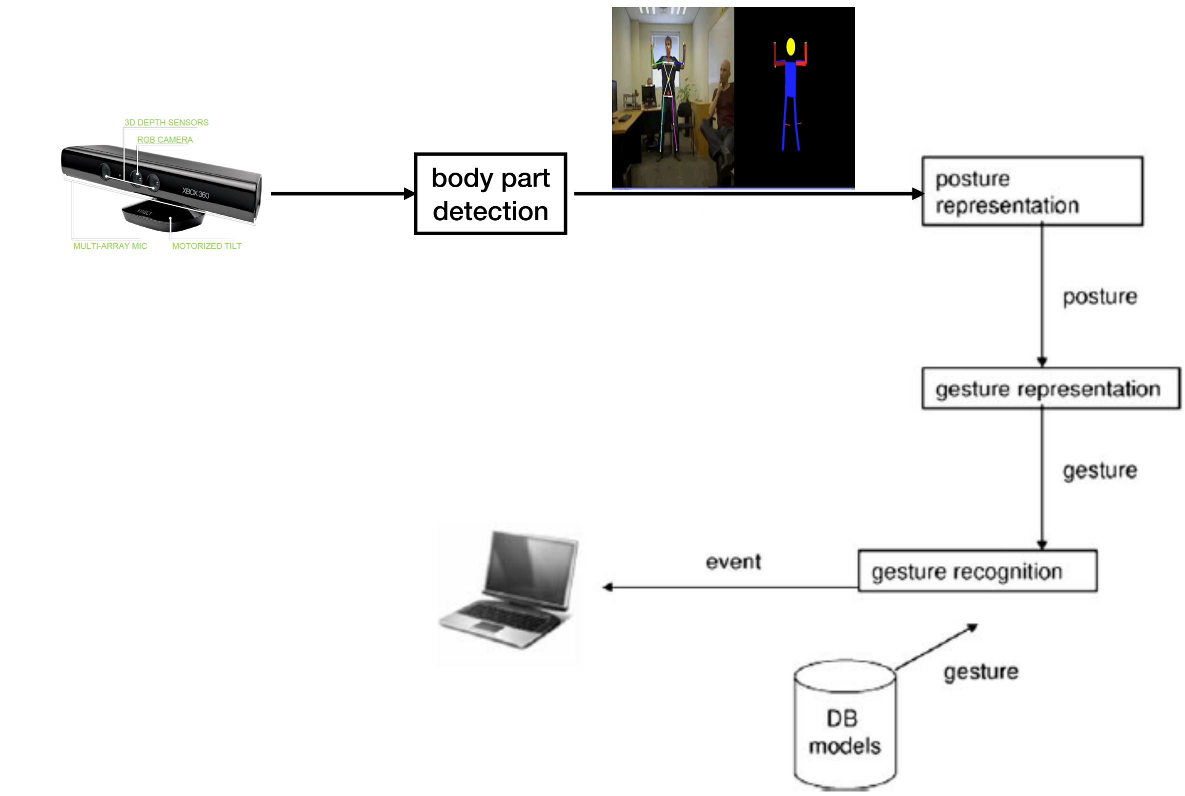
http://vision.ucsd.edu/~pdollar/toolbox/doc/index.html

Analysing postures and gestures:
//detecting parts: computer vision

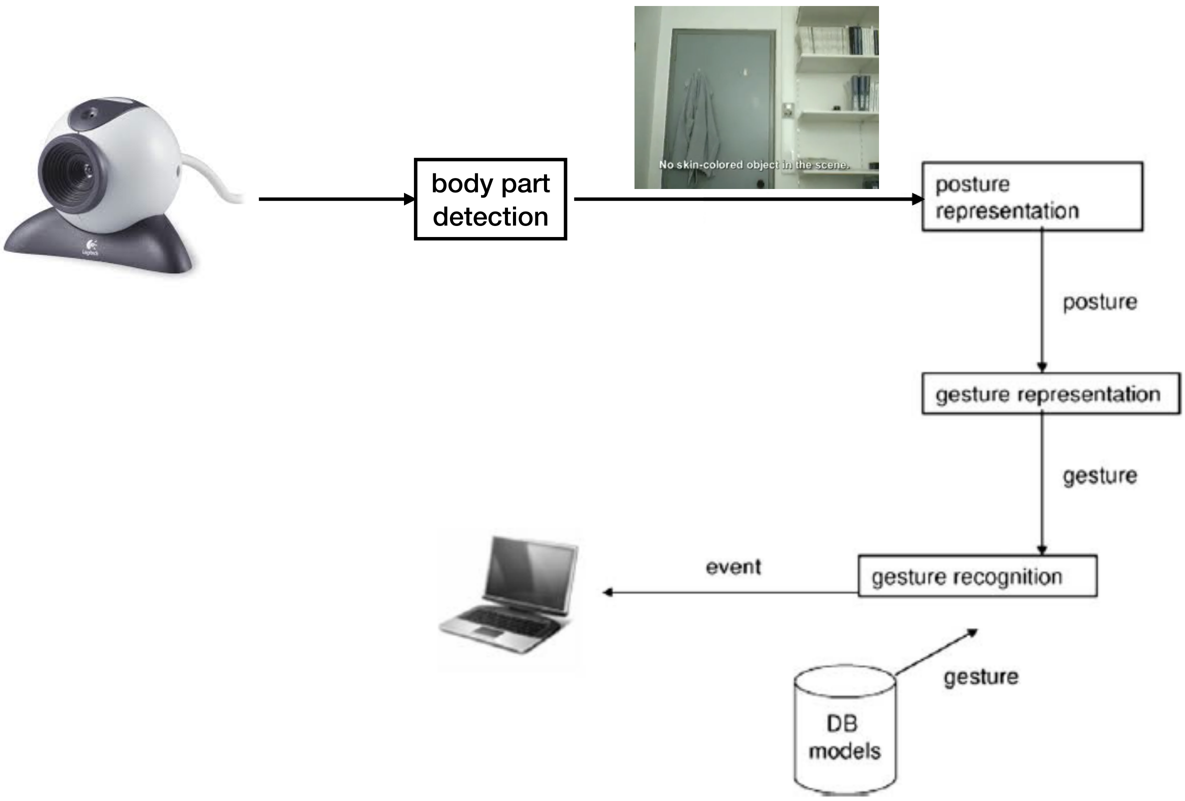




Analysing postures and gestures



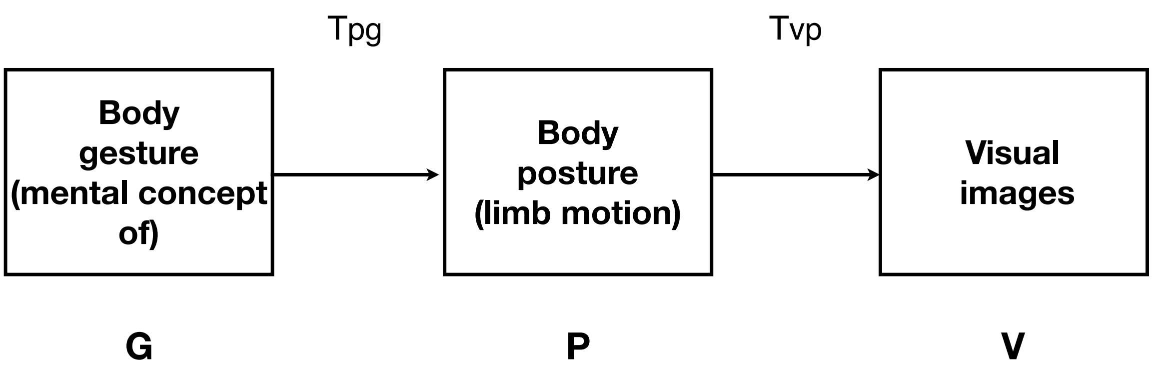
Analysing postures and gestures



Analysing postures and gestures

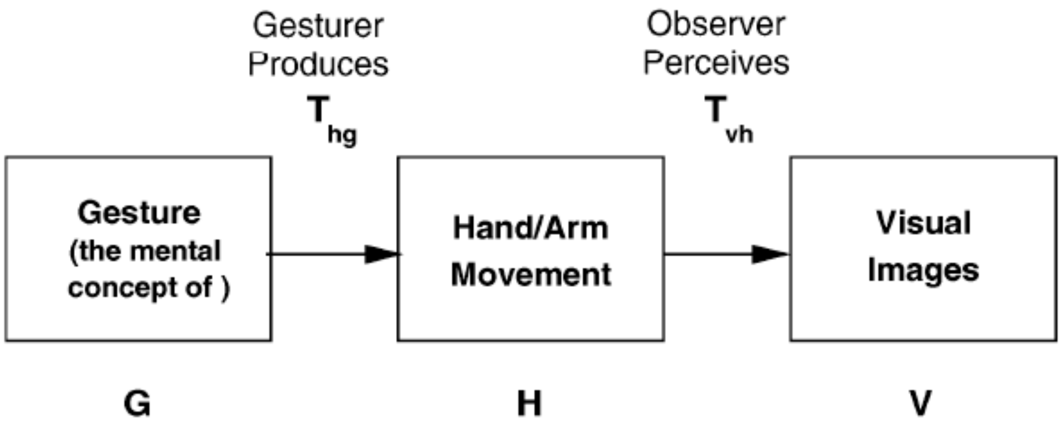
- There are two main challenges in recognizing posture and gestures:
    - detecting the body parts involved in the gesture (e.g. hands)
    - modeling the temporal dynamic of the gesture
        - addressed by using techniques such as Dynamic Time Warping , Hidden Markov Models, and Conditional Random Fields .

Body gestures and postures //Generative model



Production and perception of body gestures. Body gestures originate as a mental concept G, are expressed (Tpg) through limb motion motion B, and are perceived (Tvb) as visual images V.

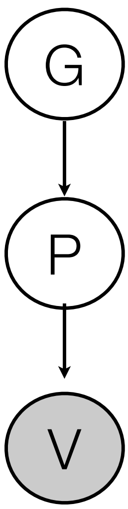
Hand gestures
//Generative model



Production and perception of gestures. Hand gestures originate as a mental concept G, are expressed (T_{hg}) through arm and hand motion H, and are perceived (T_{vh}) as visual images V.

Body gestures and postures
//Generative model

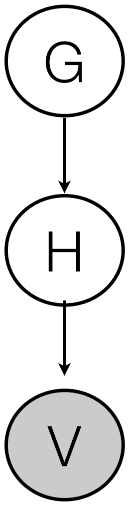
$$P(V, P, G)= P(V \mid P)\, P(P \mid G)\, P(G)$$



Production and perception of body gestures. Body gestures originate as a mental concept G, are expressed (T_{pg}) through limb motion motion B, and are perceived (T_{vb}) as visual images V.

Hand gestures
//Generative model

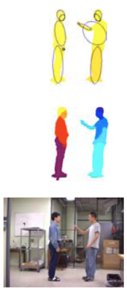
$$P(V, H, G) = P(V \mid H)\, P(H \mid G)\, P(G)$$

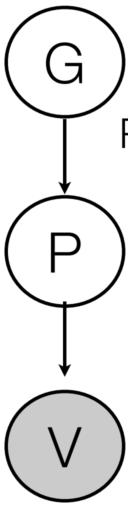


Production and perception of gestures. Hand gestures originate as a mental concept G, are expressed (Thg) through arm and hand motion H, and are perceived (Tvh) as visual images V.

Body gestures and postures
//Generative model: inference

pointing





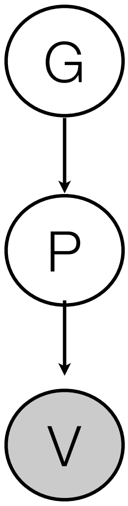
$$P(G \mid V) = \frac{P(V \mid G)\, P(G)}{P(V)}$$

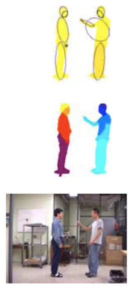
$$= \frac{P(G) \sum_H P(V \mid P)\, P(P \mid G)}{\sum_G \sum_H P(G)\, P(V \mid P)\, P(P \mid G)}$$

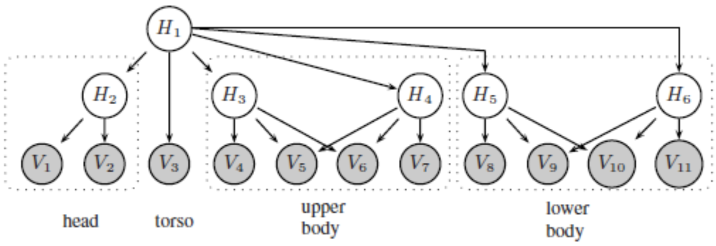
Production and **perception** of body gestures. Body gestures originate as a mental concept G, are expressed (Tpg) through limb motion motion B, and are perceived (Tvb) as visual images V.

Body gestures and postures
//Generative model: more complex model

pointing

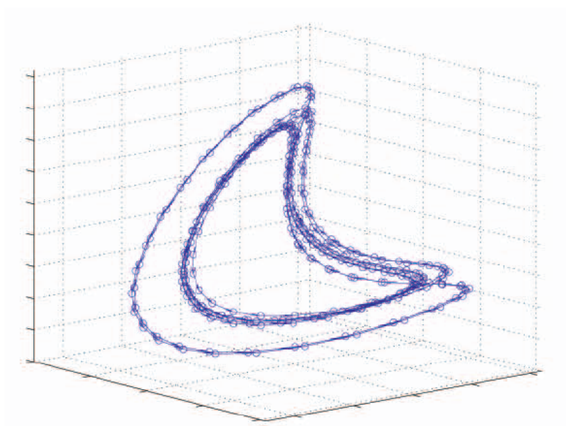




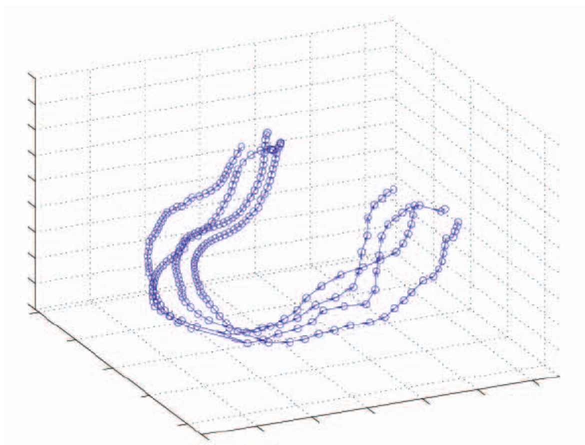


Production and **perception** of body gestures. Body gestures originate as a mental concept G, are expressed (Tpg) through limb motion motion B, and are perceived (Tvb) as visual images V.

Body gestures and postures
//Generative model



walking



golf swing

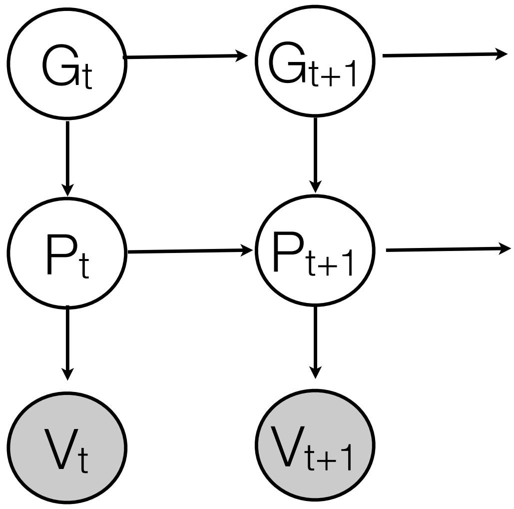
A ***body*** *gesture is a stochastic process in the gesture model parameter space $\mathcal{M}_T$ over a suitably defined time interval I.*

Body gestures and postures
//Generative model: inference

$P(G_{t+1} \mid V_{t+1}) \approx$

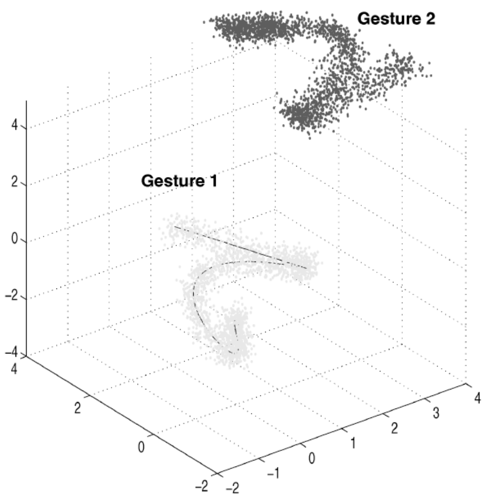
$P(V_{t+1} \mid G_{t+1})\, P(G_{t+1} \mid V_t) \approx$

$P(V_{t+1} \mid G_{t+1}) \sum_{G_t} P(G_{t+1} \mid G_t)\, P(G_t \mid V_t)$



*A **body** gesture is a stochastic process in the gesture model parameter space $\mathcal{M}_T$ over a suitably defined time interval I.*

Hand gestures
//Generative model



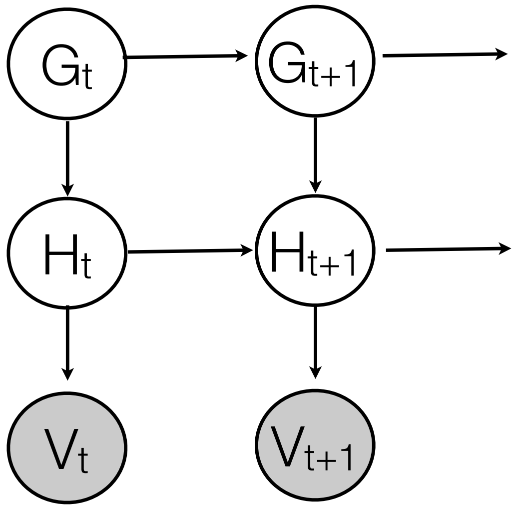
A hand gesture is a stochastic process in the gesture model parameter space $\mathcal{M}_T$ over a suitably defined time interval I.

Hand gestures
//Generative model

$$P(G_{t+1} \mid V_{t+1}) \approx$$

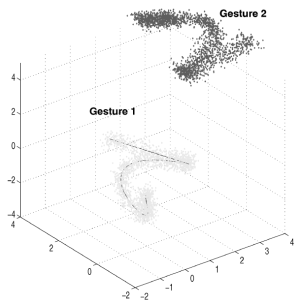
$$P(V_{t+1} \mid G_{t+1})\, P(G_{t+1} \mid V_t) \approx$$

$$P(V_{t+1} \mid G_{t+1}) \sum_{G_t} P(G_{t+1} \mid G_t)\, P(G_t \mid V_t)$$



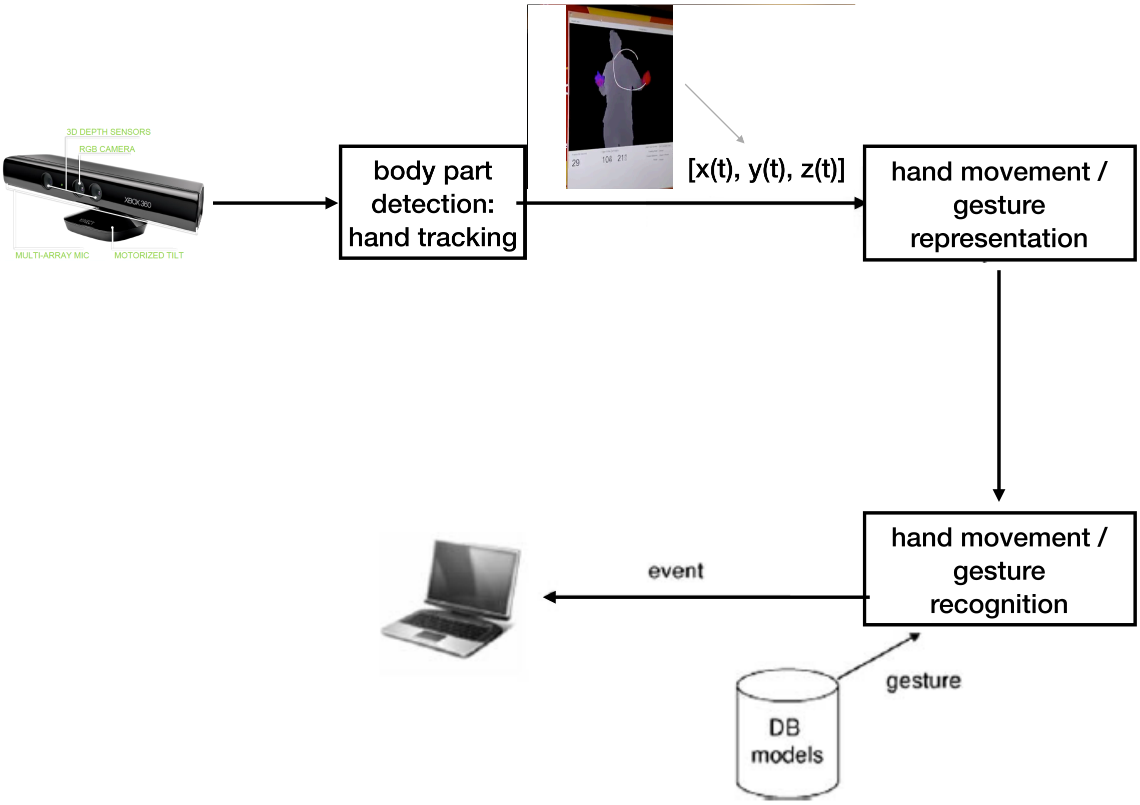
A hand gesture is a stochastic process in the gesture model parameter space $\mathcal{M}_T$ over a suitably defined time interval $\mathcal{I}$.

Hand gesture recognition
//Simple example

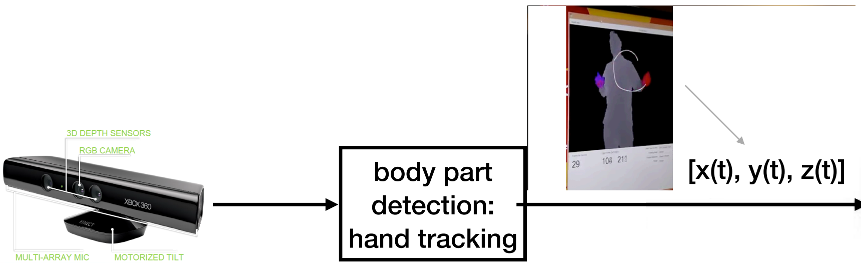


A hand gesture is a stochastic process in the gesture model parameter space $\mathcal{M}_T$ over a suitably defined time interval $\mathcal{I}$.

Hand gesture recognition
//Simple example: app architecture



Hand gesture recognition
//Simple example: app architecture



Features: sequential 3D points (x, y, z) of the hand

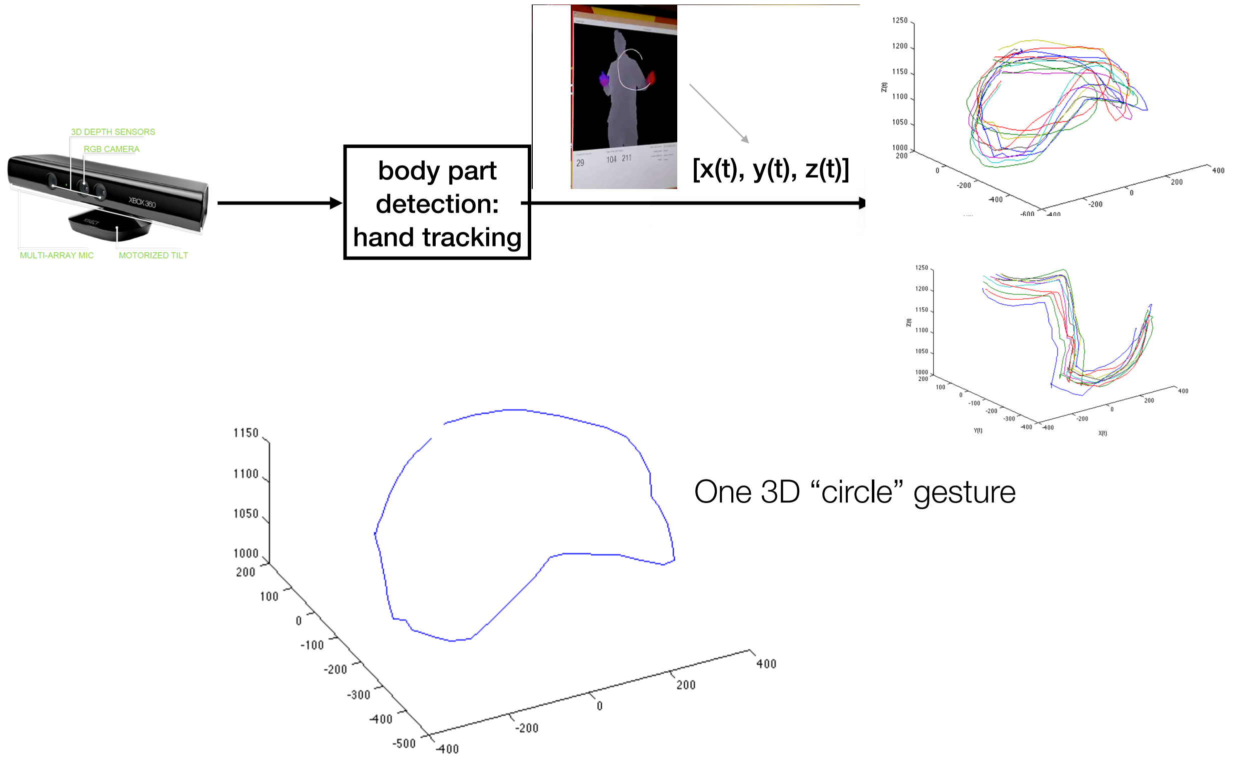
- Computed from the hand-point tracking code implemented in PrimeSense open-source OpenNI framework for Kinect

- Each gesture repeated 10 times

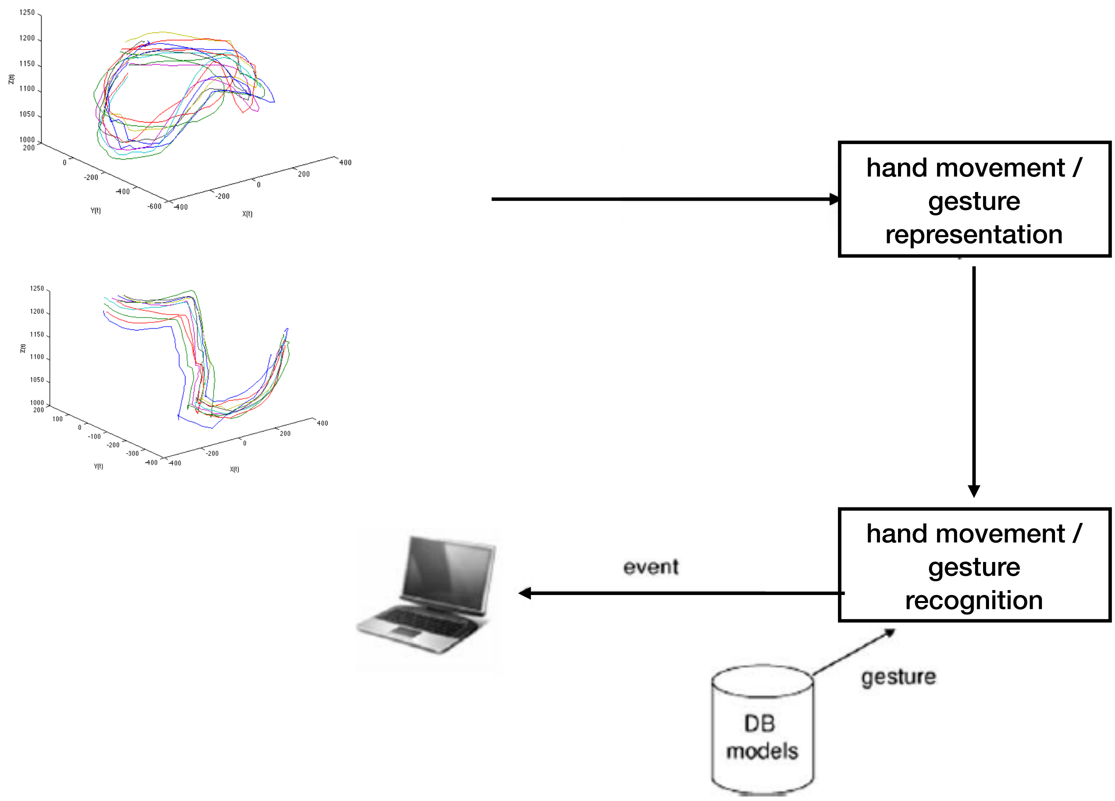
- OpenNI hand-tracking code runs at 30 frames per second,
a two-second circle gesture is captured in 60 observations (trajectory sampled points),

- Each observation having three dimensions (x, y, z)

Hand gesture recognition
//Simple example: app architecture



Hand gesture recognition
//Simple example: app architecture

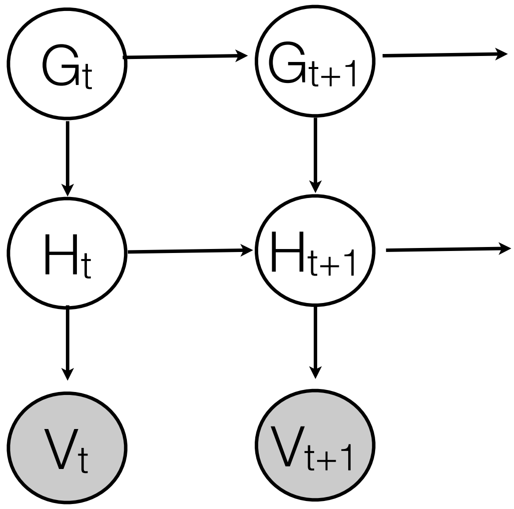


Hand gesture recognition
//Simple example: gesture inference

$$P(G_{t+1} \mid V_{t+1}) \approx$$

$$P(V_{t+1} \mid G_{t+1})\, P(G_{t+1} \mid V_t) \approx$$

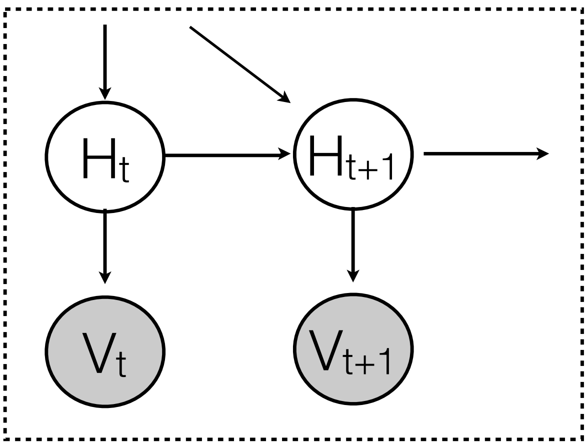
$$P(V_{t+1} \mid G_{t+1}) \sum_{G_t} P(G_{t+1} \mid G_t)\, P(G_t \mid V_t)$$



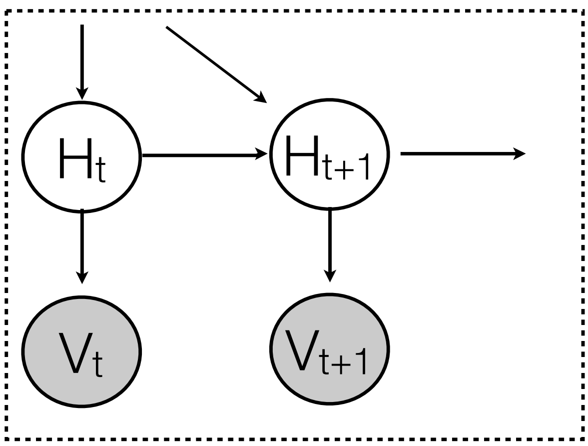
hand movement / gesture recognition

Hand gesture recognition
//Simple example: gesture inference

G=circle



G=other



Each gesture is modelled by a specific HMM
Can compute the log-likelihood of the hand movement, given the gesture:

$$\log P(H_{t+1}, H_t, H_{t-1}, \ldots \mid G = \text{circle})$$

Hand gesture recognition
//Simple example: gesture inference

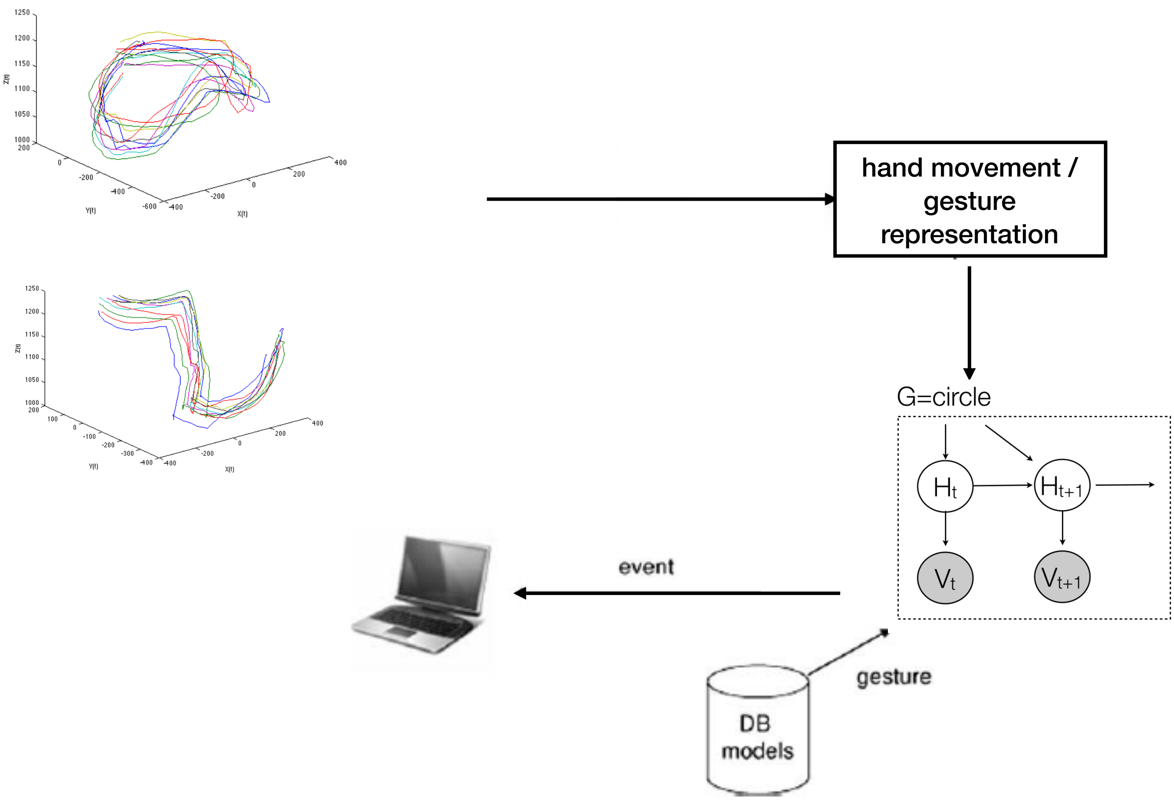
Prior choice: restrict the model to the sequence of adjacent states that constitute a gesture.

- zeroing out the probabilities at the matrix edges: the probabilities of jumping between distant states in the model

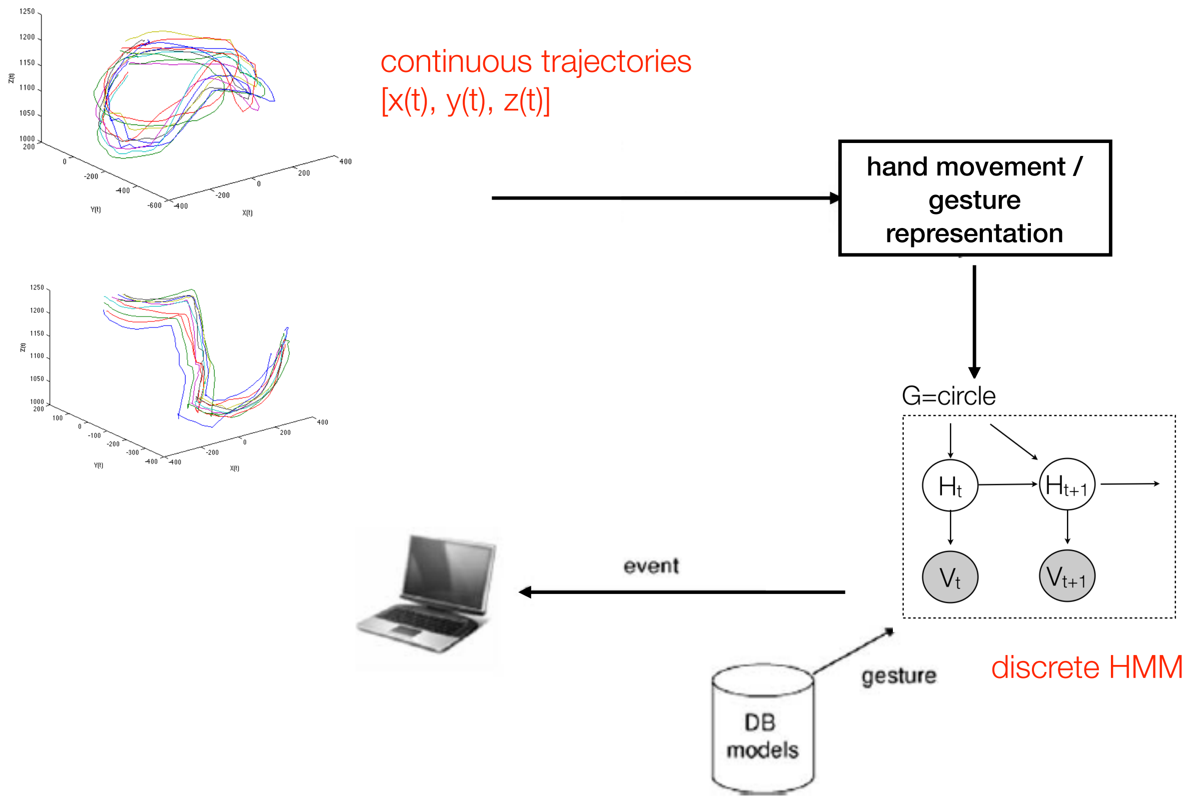
- choose a prior that constrains the transition matrix such that state transitions can only occur in one direction and between two adjacent states: a left-to-right HMM

$$A = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

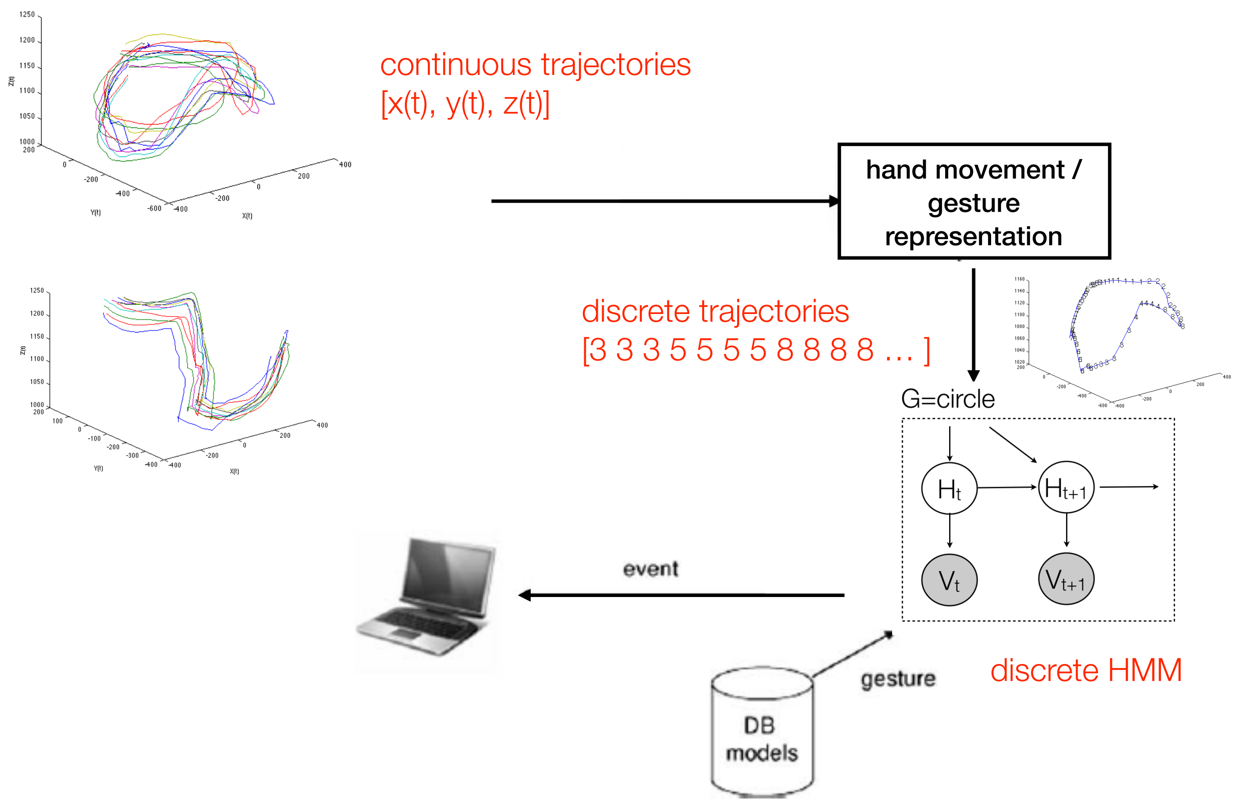
Hand gesture recognition
//Simple example: app architecture



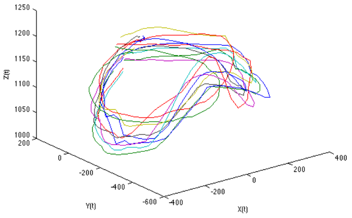
Hand gesture recognition
//Simple example: app architecture



Hand gesture recognition
//Simple example: app architecture

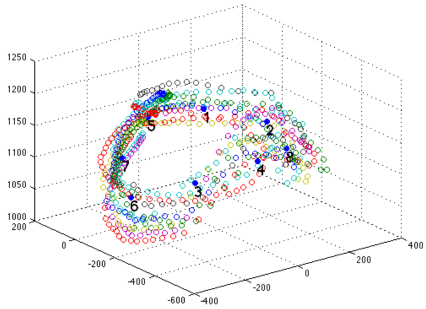


Hand gesture recognition
//Simple example: app architecture

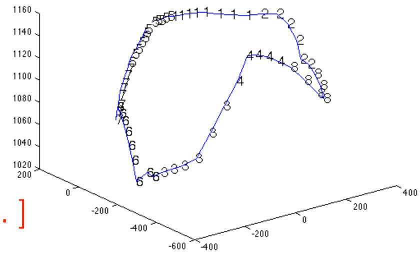


continuous trajectories
[x(t), y(t), z(t)]

k-means point clustering of 10 gesture repetitions

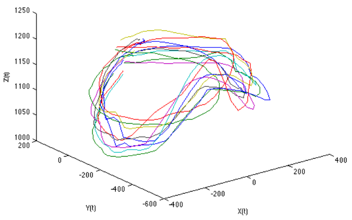


centroid-based quantization of trajectory



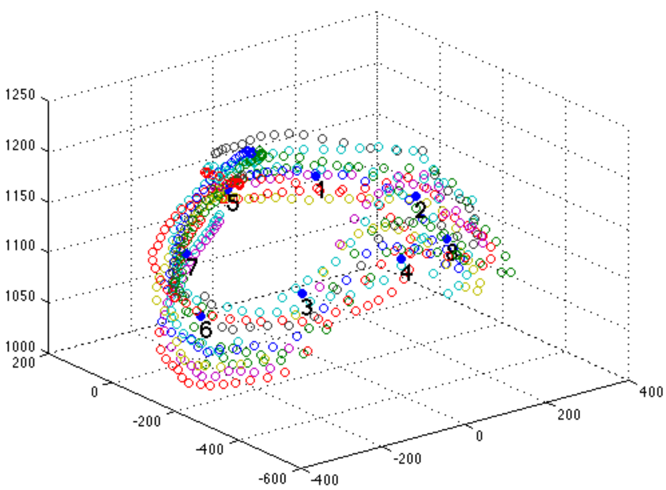
discrete trajectories
[3 3 3 5 5 5 5 8 8 8 8 …]

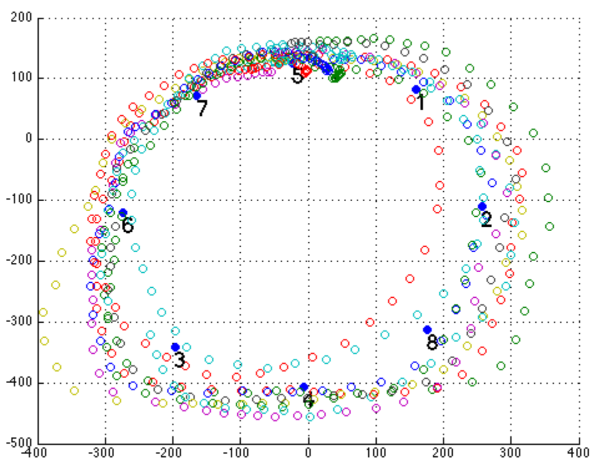
Hand gesture recognition
//Simple example: app architecture



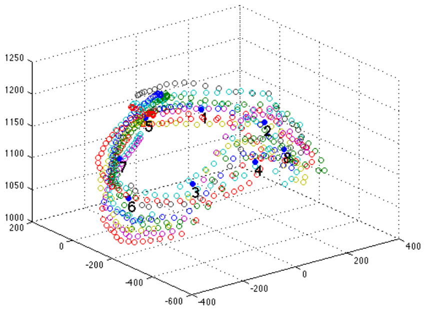
continuous trajectories
[x(t), y(t), z(t)]

k-means point clustering of 10 gesture repetitions

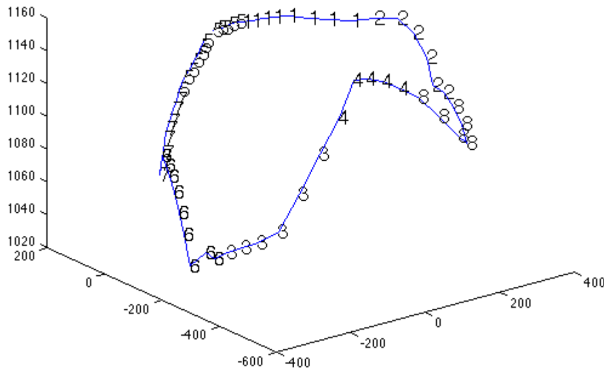


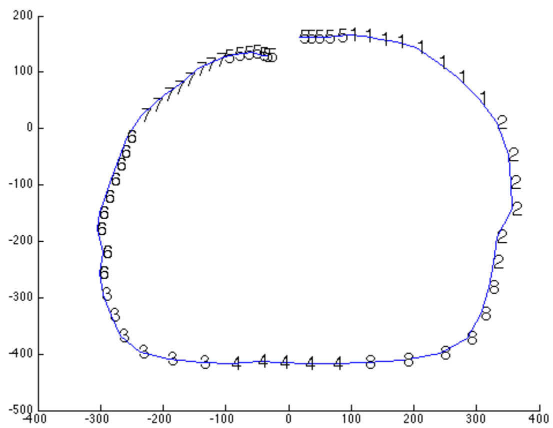


Hand gesture recognition
//Simple example: app architecture

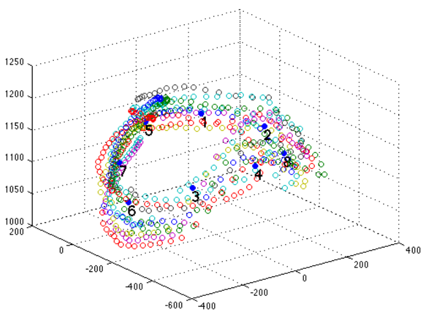


centroid-based quantization of trajectory

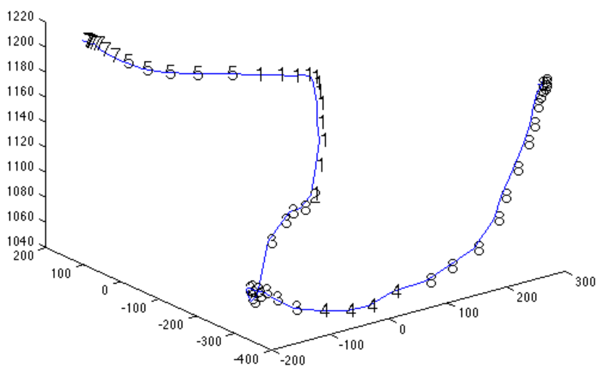


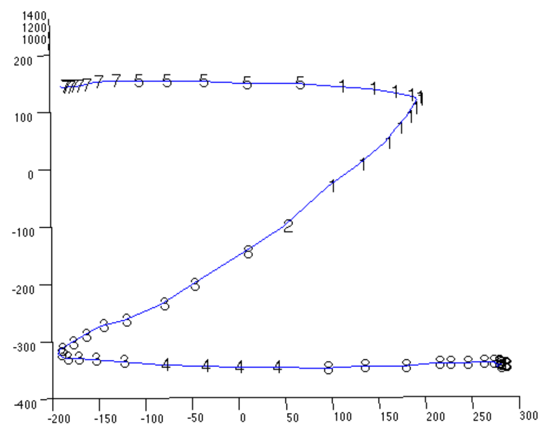


Hand gesture recognition
//Simple example: app architecture



centroid-based quantization of trajectory





Hand gesture recognition
//Simple example: app architecture