Activity Recognition and 3D Lower Limb Tracking with an Instrumented Walker

Wednesday, April 6, 2011
NICTA, Canberra

Pascal Poupart
Associate Professor
Cheriton School of Computer Science
University of Waterloo

Smart Walker Project

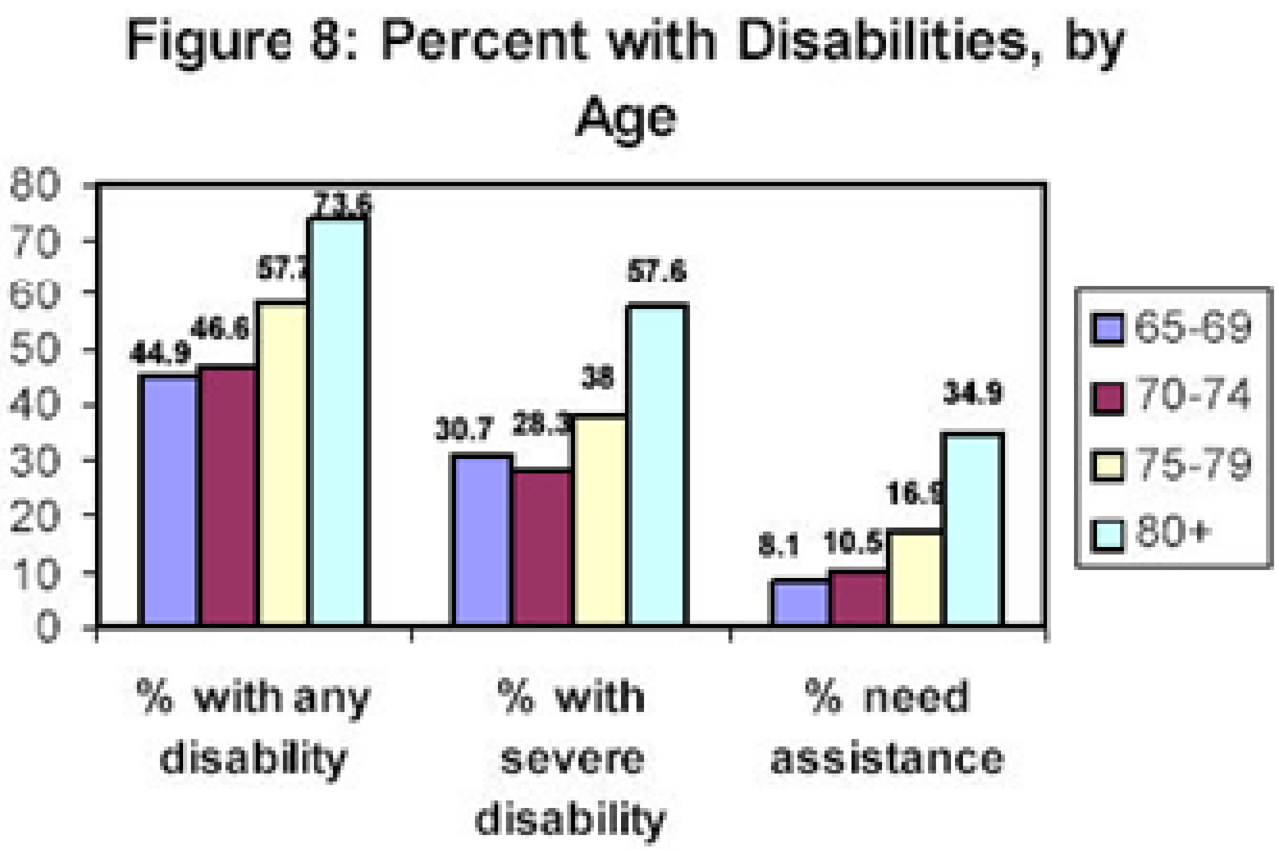
- Smart walker: rollating walker instrumented with sensors
- Problems:
  - Activity recognition
  - Lower limb 3D pose tracking
- Contributions:
  - HMM and CRF modeling
  - Comparative analysis of algorithms for
    - Activity recognition
    - 3D pose tracking
  - Experimentation with regular walker users

Outline

- Motivation
- Activity recognition
  - Experimental data
  - Algorithms (supervised/unsupervised HMMs and CRFs)
  - Results
- 3D lower limb tracking
  - Challenges
  - Trackers (RGB binocular and structured light cameras)
  - Results
- Conclusion & future work

Disability Statistics

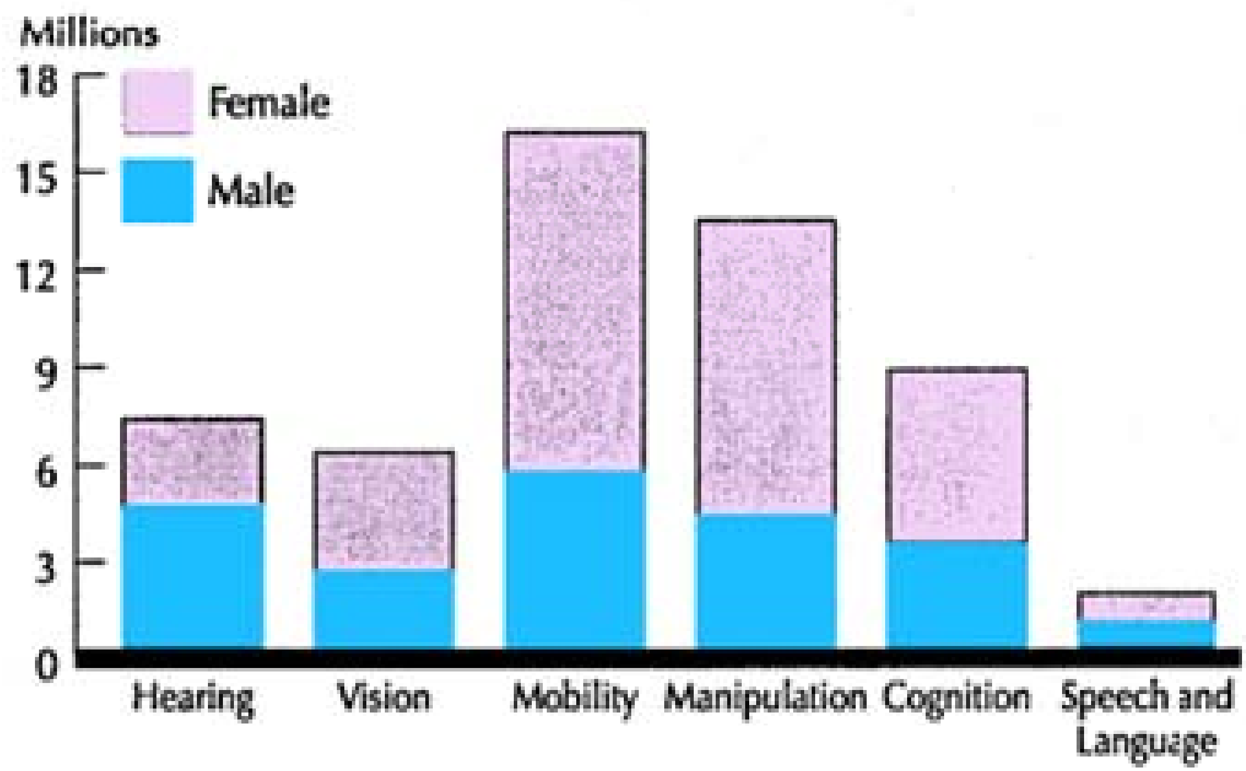
- US National Business Service Alliance (July 2006)



Disability Statistics

- US National Business Service Alliance (July 2006)

Types of disabilities



pp3

Mobility aids

Encourage walking
Increase safety

Discourage walking

pp3 What are the walking aids available to seniors?
Pascal Poupart, 11/21/2008

Smart Walker

+

Force sensors
Load sensors
Video cameras
Microphone
Speech synthesizer
Servo-brakes
etc.

walker

devices

Smart walker

assist

caregivers

users

Research Goals

- Prototype designed by J. Tung, B. McIlroy and K. Habib (U of Waterloo and Toronto Rehab Institute)
- Clinical assessment of walking abilities
  - Toronto Rehab and Village of Winston Park
  - Walking course with instrumented walker
    - Identify context and triggers of falls
    - Assess balance control and stability
- Goals:
  - Automated activity recognition (context)
  - 3D pose modeling (balance assessment)

Outline

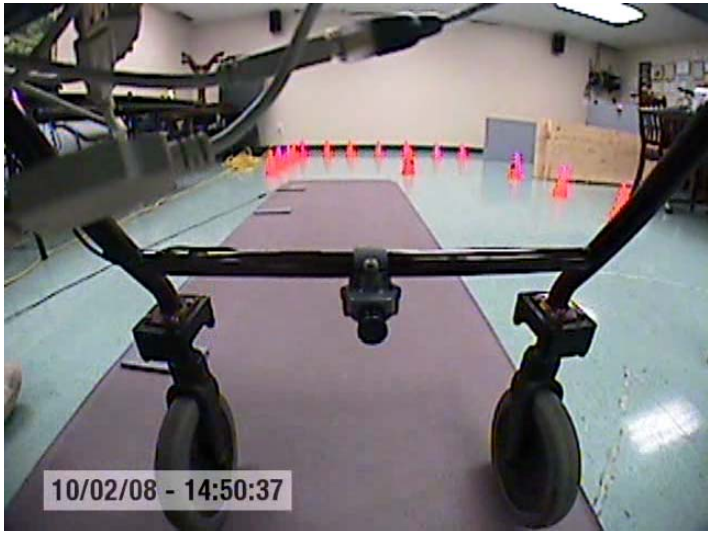
- Motivation
- Activity recognition
  - Experimental data
  - Algorithms (supervised/unsupervised HMMs and CRFs)
  - Results
- 3D lower limb tracking
  - Challenges
  - Trackers (RGB binocular and structured light cameras)
  - Results
- Conclusion & future work

Activity Recognition

- State of the art: kinesiologists hand label sensor data by looking at video feeds
  - Time consuming and error prone!

Backward view

Forward view



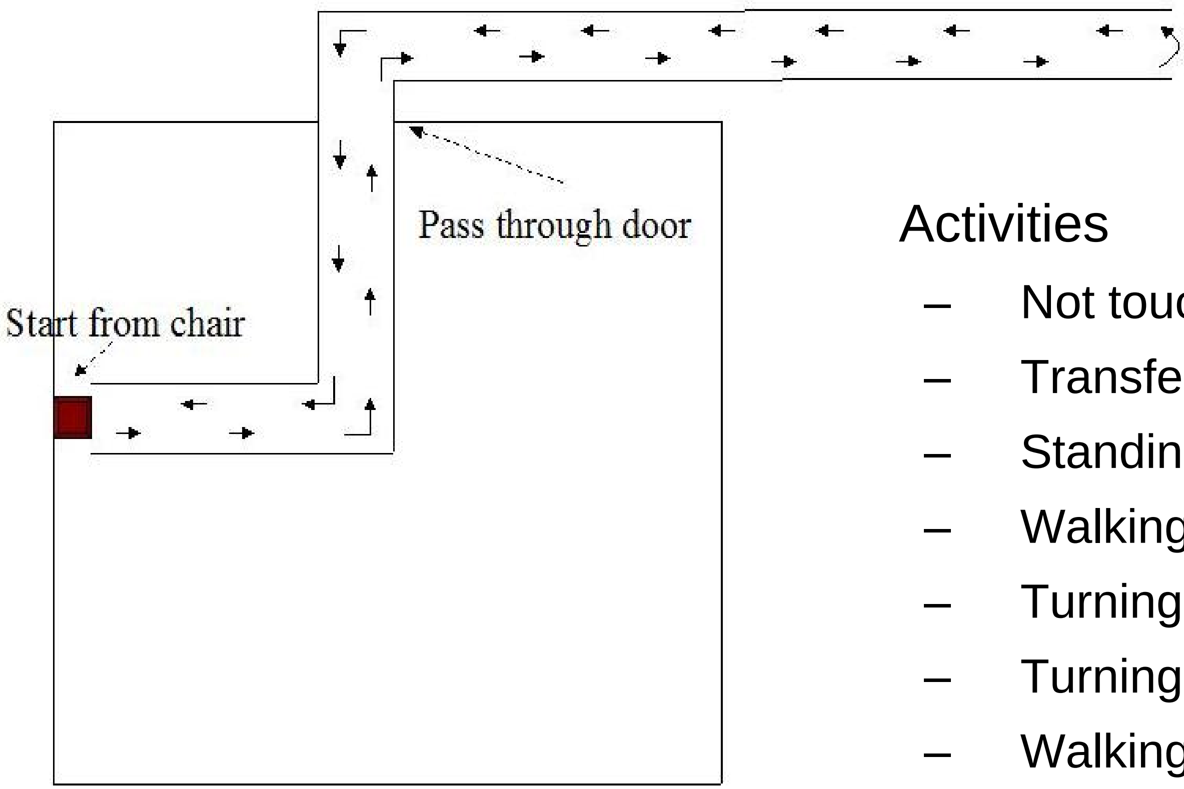
Raw Sensor Data

- 8 channels:
  - Forward acceleration
  - Lateral acceleration
  - Vertical acceleration
  - Load on left rear wheel
  - Load on right rear wheel
  - Load on left front wheel
  - Load on right front wheel
  - Wheel rotation counts (speed)

- Data recorded at 50 Hz and digitized (16 bits)

Experiment 1

- 17 healthy young subjects (19-53 years old)
- Each person performed the course twice

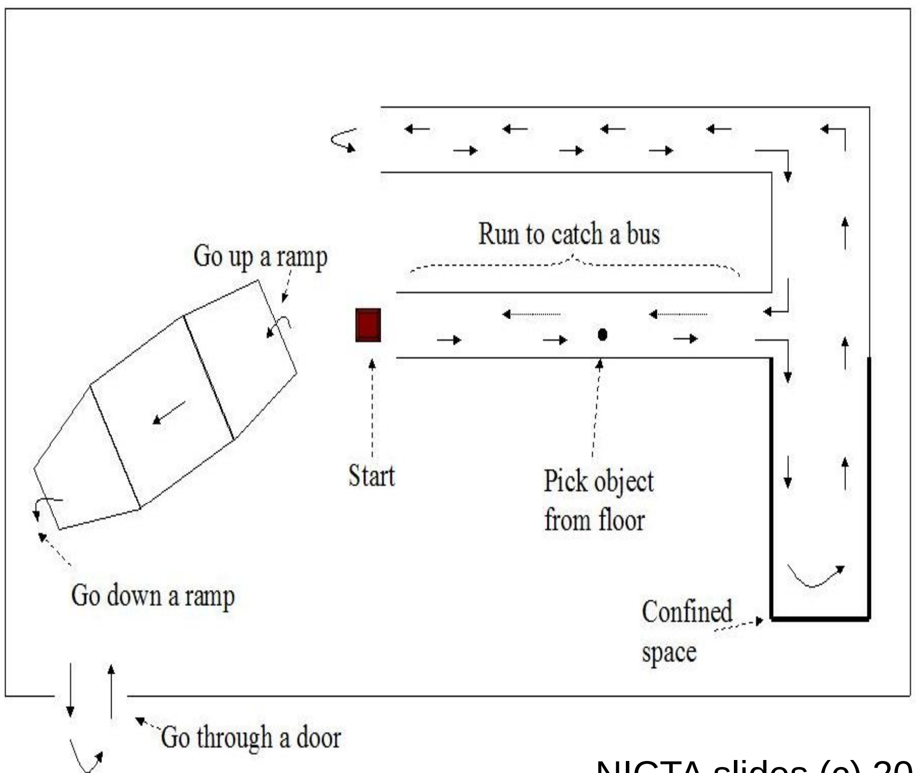


Activities

- Not touching the walker (NTW)
- Transfers (sit ↔ stand) (TR)
- Standing (ST)
- Walking Forward (WF)
- Turning Left (TL)
- Turning Right (TR)
- Walking Backwards (WB)

Experiment 2

- 8 walker users at Winston Park (84-97 years old)
- 12 older adults (80-89 years old) in the Kitchener-Waterloo area who do not use walkers

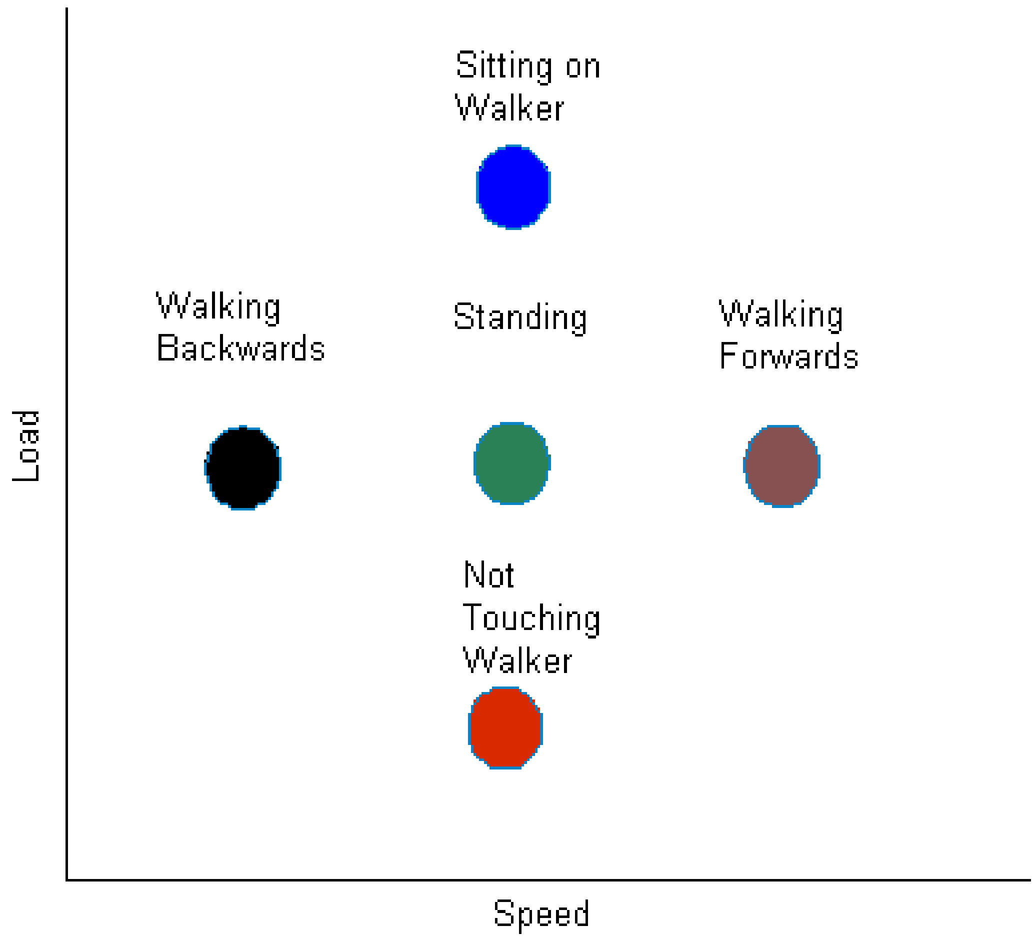


Activities

- Not Touching Walker (NTW)
- Standing (ST)
- Walking Forward (WF)
- Turning Left (TL)
- Turning Right (TR)
- Walking Backwards (WB)
- Sitting on the Walker (SW)
- Reaching Tasks (RT)
- Up Ramp/Curb (UR/UC)
- Down Ramp/Curb (DR/DC)

Hypotheses I

- Activities expected to be easily distinguishable



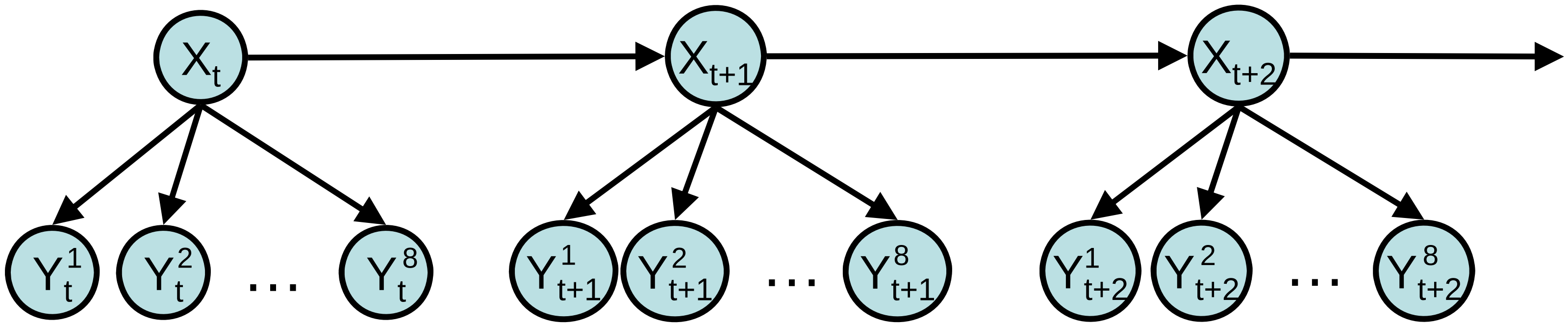
Hypotheses II

- Less obvious distinctions due to a variety of behaviours
  - Turns:
    - load fluctuations on side of turn
    - reduced speed compared to walking
    - Mild lateral force
  - Vertical transitions:
    - Fluctuations in vertical acceleration
    - Load fluctuations
  - Transfers:
    - Fluctuations in rear load
    - Mild movements
  - Reaching tasks
    - Unclear effect on sensors

Probabilistic Models

- Hidden Markov Model (HMM)
  - Supervised
    - Maximum likelihood (ML)
  - Unsupervised
    - Expectation maximization (EM)
    - Bayesian Learning
- Conditional Random Field (CRF)
  - Supervised
    - Maximum conditional likelihood
    - Automated feature extraction

Hidden Markov Model (HMM)



- Parameters
  - Initial state distribution: $\pi_x = \Pr(X_0 = x)$
  - Transition probabilities: $\theta_{x'|x} = \Pr(X_{t+1} = x' | X_t = x)$
  - Emission probabilities: $\phi^i_{y|x} = \Pr(Y^i_t = y | X_t = x)$

Supervised Maximum Likelihood

- Supervised learning:
  - Relative frequency counts

$$\pi_x \leftarrow \frac{\#x}{n} \qquad \theta_{x'|x} \leftarrow \frac{\#<x',x>}{\#x} \qquad \phi_{y|x} \leftarrow \frac{\#<y,x>}{\#x}$$

  - Alternatively, to avoid bias due to predefined course

$$\pi \leftarrow uniform \qquad \theta_{x'|x} \leftarrow \begin{cases} p & \text{if } x' = x \\ \frac{1-p}{|\mathcal{X}|-1} & \text{otherwise} \end{cases}$$

Unsupervised Maximum Likelihood

- Expectation maximization (EM)
  - Expectations

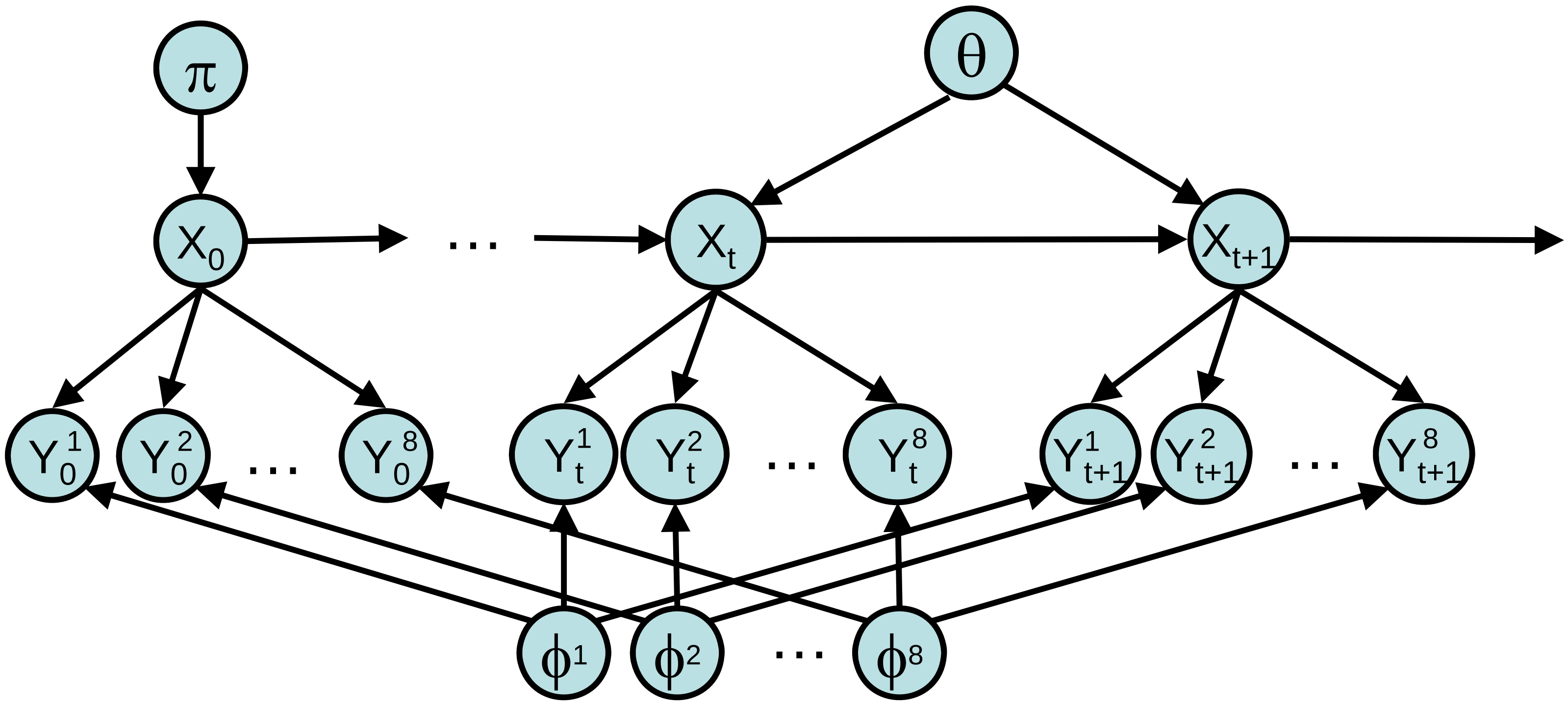
$$E_0(x) = \Pr(X_0 = x | y_{1..T}, \pi, \theta, \phi)$$
$$E(x, x') = \sum_t \Pr(X_t = x, X_{t+1} = x' | y_{1..T}, \pi, \theta, \phi)$$
$$E(x, y) = \sum_t \Pr(X_t = x, Y_t = y | y_{1..T}, \pi, \theta, \phi)$$

  - Improvement

$$\pi_x \leftarrow E_0(x)$$
$$\theta_{x'|x} \leftarrow \frac{E(x,x')}{\sum_{x'} E(x,x')}$$
$$\phi_{y|x} \leftarrow \frac{E(x,y)}{\sum_y E(x,y)}$$

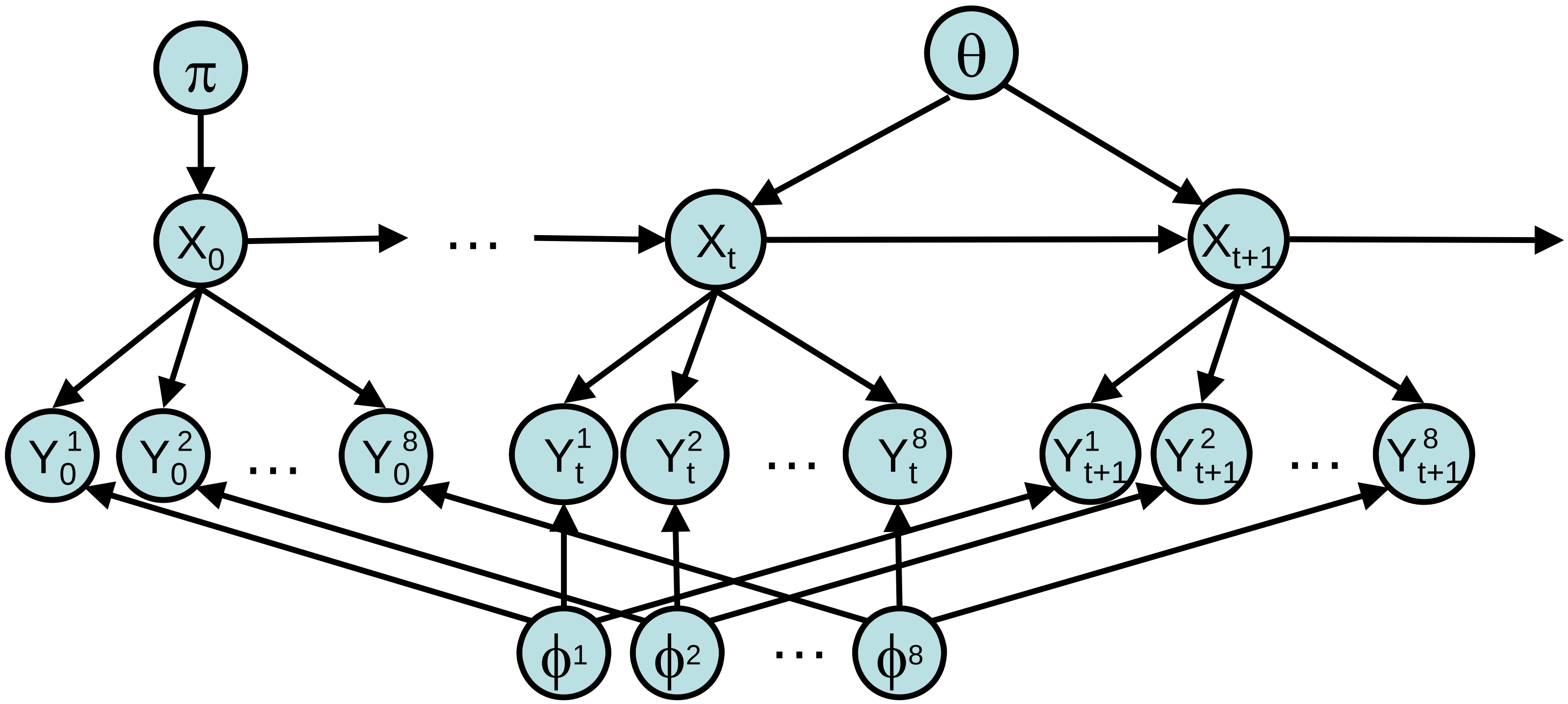
- May overfit or get stuck in local optima!

Bayesian Learning



- Parameters treated as random variables
- More robust to overfitting and local optima

Bayesian Learning



Conditional distributions

$$\Pr(X_0 = x|\pi) = \pi_x$$

$$\Pr(X_{t+1} = x'|X_t = x, \theta) = \theta_{x'|x}$$

$$\Pr(Y_t^i = y|X_t = x, \phi) = \phi^i_{y|x}$$

Priors

$$\pi_{\cdot} \sim \text{Dirichlet}$$

$$\theta_{\cdot|x} \sim \text{Dirichlet}$$

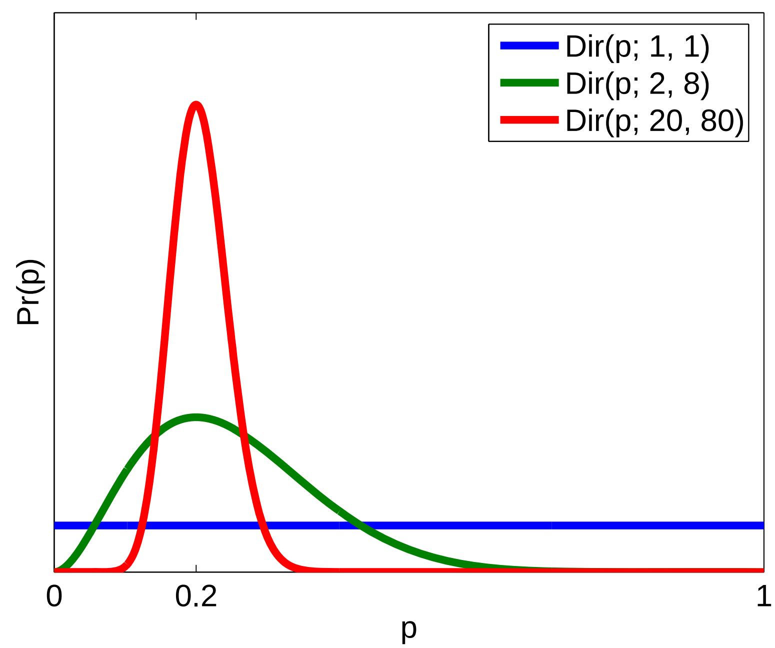
$$\phi^i_{\cdot|x} \sim \text{Dirichlet}$$

Dirichlet Distributions

- Dirichlets are monomials over discrete random variables

$$Dir(p; n) = k \prod_i p_i^{n_i - 1}$$

- Properties of Dirichlets
  - Easy to integrate
  - Conjugate prior for discrete likelihood distributions



Bayesian Learning

- Parameter learning: compute posterior

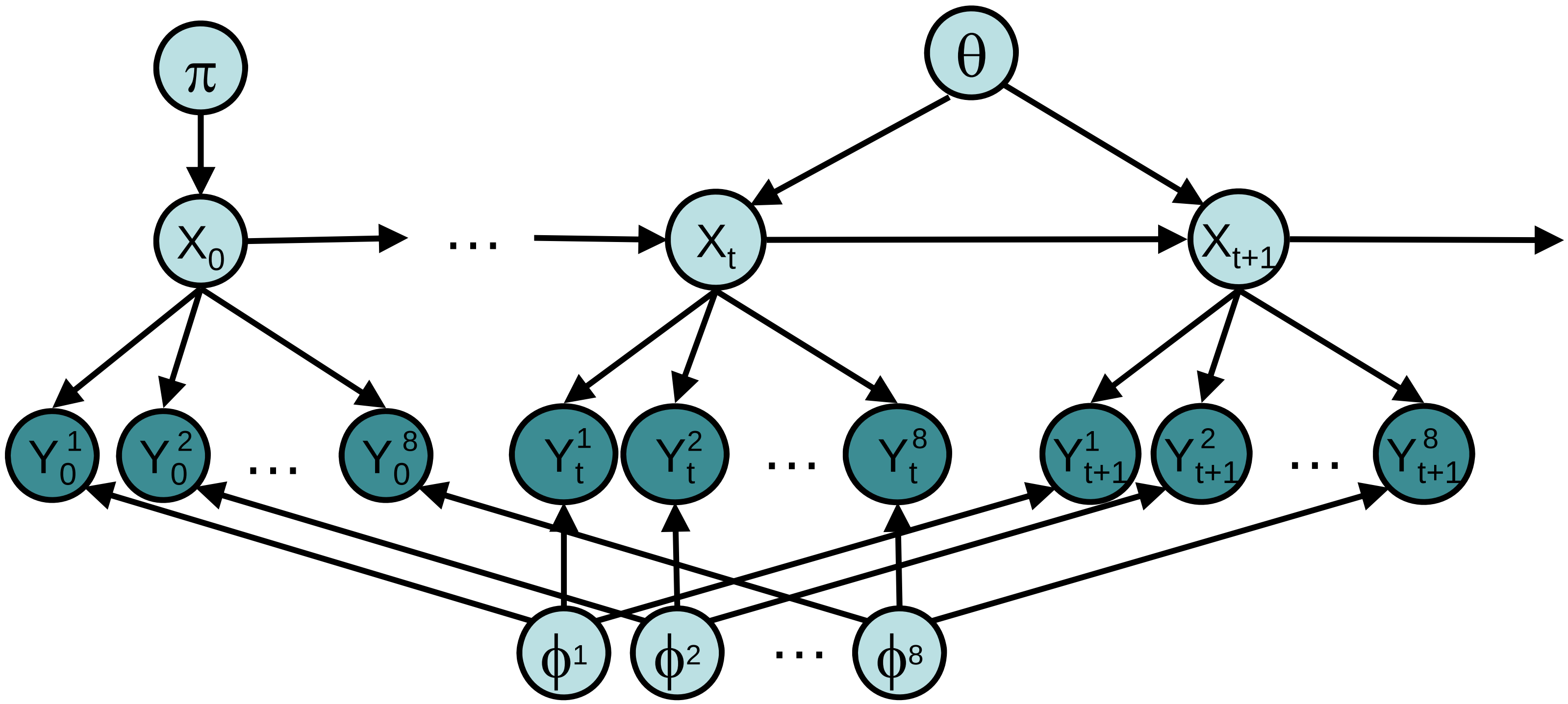
$$\Pr(\pi,\theta,\phi|y_{0..T}) \propto \Pr(\pi)\Pr(\theta)\Pr(\phi)\Pr(y_{0..T}|\pi,\theta,\phi)$$

- Intractable: exponential mixture of Dirichlets

$$\begin{aligned}
&\Pr(\pi,\theta,\phi|y_{0..T}) \\
&\quad \propto \Pr(\pi)\Pr(\theta)\Pr(\phi)\textstyle\sum_{x_{0..T}}\Pr(x_0|\pi)\prod_t \Pr(x_{t+1}|x_t,\theta)\Pr(y_t|x_t,\phi) \\
&\quad = mono(\pi)mono(\theta)mono(\phi)\textstyle\sum_{x_{0..T}}\pi_{x_0}\prod_t \theta_{x_{t+1}|x_t}\phi_{y_t|x_t} \\
&\quad = polynomial(\pi,\theta,\phi)
\end{aligned}$$

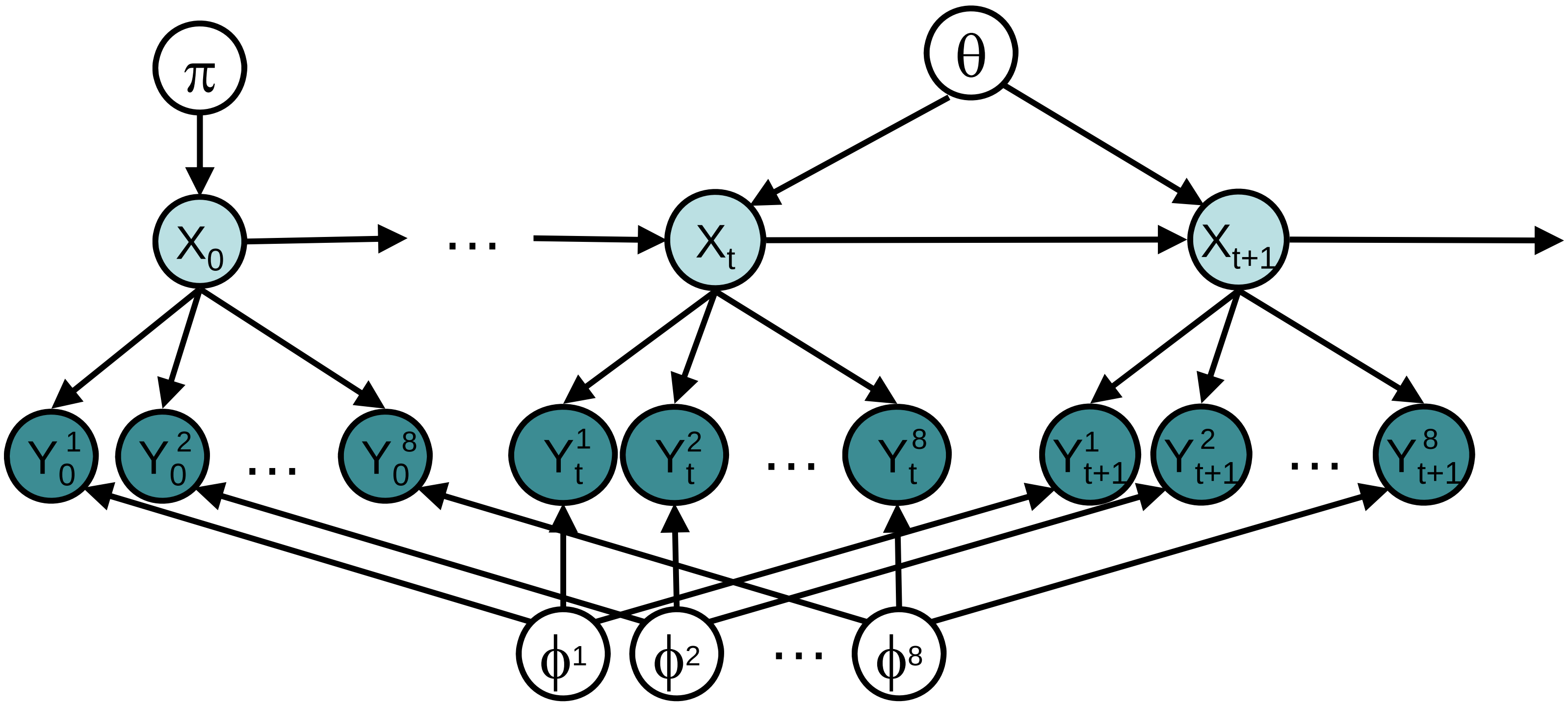
- Approximate solution: Gibbs sampling

Gibbs Sampling



- Markov chain that converges to the posterior
- Repeatedly sample $X_{0..T}, \pi, \theta, \phi$

Collapsed Gibbs Sampling



- Faster convergence
- Repeatedly sample $X_{0..T}$ while integrating out π, θ, ϕ

Prediction

- Maximum aposteriori filtering

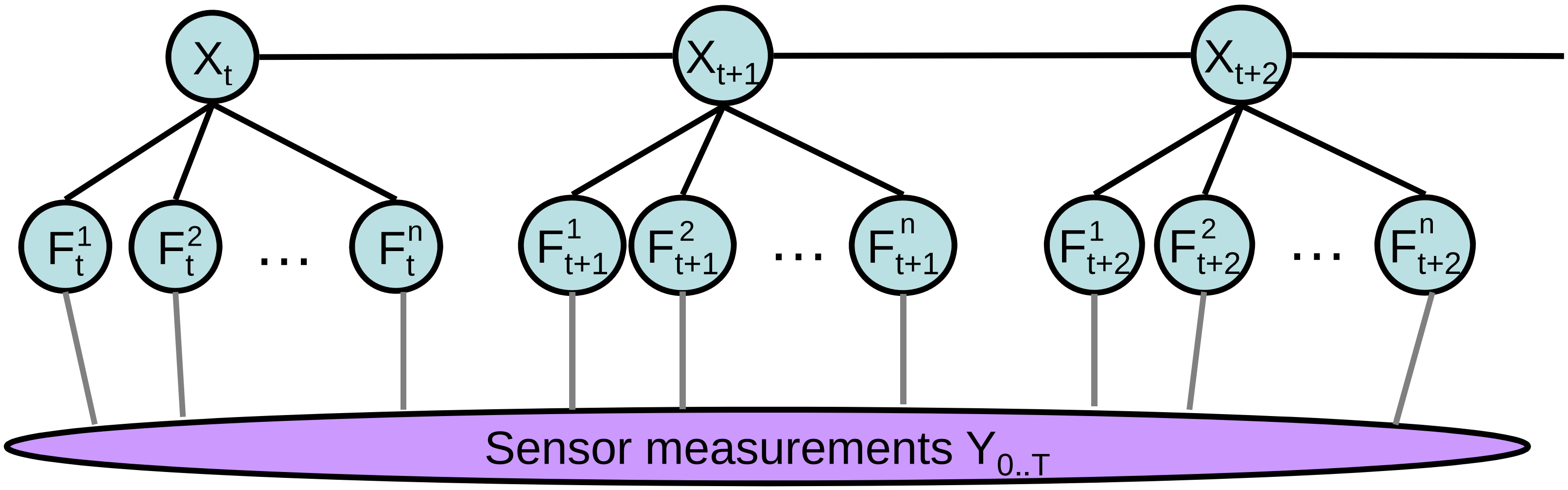
$$x_t^* = \mathrm{argmax}_x \Pr(X_t = x | y_{0..t}, \pi, \theta, \phi)$$

- Recursive belief monitoring

$$\begin{aligned} &\Pr(X_t | y_{0..t}, \pi, \theta, \phi) \\ &\quad \propto \textstyle\sum_{x_{t-1}} \Pr(X_t | x_{t-1}, \theta) \Pr(y_t | X_t, \phi) \Pr(x_{t-1} | y_{0..t-1}, \pi, \theta, \phi) \end{aligned}$$

- Online prediction in real-time

Conditional Random Field (CRF)



- Discriminative training: maximum conditional likelihood
- Does not assume conditional independence of sensor measurements
- Use features instead of raw measurements

Conditional Random Field (CRF)

- Probabilistic model

$$\Pr(X_{0..T} = x_{0..T} | Y_{0..T} = y_{0..T}) \propto e^{\sum_i \lambda_i f_i(x_{0..T}, y_{0..T})}$$

- Transition features: capture persistence

$$f(x_t, x_{t+1}) = \begin{cases} 1 & \text{if } x_t = x_{t+1} \\ 0 & \text{otherwise} \end{cases}$$

Conditional Random Field (CRF)

- Sensor features: thresholded average and std dev
  - Let $g(y_{t\text{-}w..t})$ be the average or std dev of
    - speed, total load, center of pressure (media-lateral, anterior-posterior), acceleration (forward, lateral, vertical)
    - For windows of 1, 5, 25 and 50 measurements
  - Features take the form

$$f^{\leq}_{x,x'}(X_t, y_{t-w..t}) = \begin{cases} 1 & \text{if } X_t = x \text{ and } g(y_{t-w..t}) \leq \tau_{x,x'} \\ 0 & \text{otherwise} \end{cases}$$

$$f^{\geq}_{x,x'}(X_t, y_{t-w..t}) = \begin{cases} 1 & \text{if } X_t = x' \text{ and } g(y_{t-w..t}) \geq \tau_{x,x'} \\ 0 & \text{otherwise} \end{cases}$$

  - Thresholds τ
    - Initially: handcrafted based on data inspection
    - Later: learned by linear regression and logistic regression

Automated feature extraction

- Linear discriminant analysis:

$$\tau_{x,x'} = \frac{1}{2}\sum_t \frac{1(x_t{=}x)g(y_{t-w..t})}{\sum_t 1(x_t{=}x)} + \frac{1}{2}\sum_t \frac{1(x_t{=}x')g(y_{t-w..t})}{\sum_t 1(x_t{=}x')}$$

- Logistic regression

$$f^{\leq}_{x,x'}(X_t, y_{t-w..t}) = sigmoid(\gamma_{x,x'}(g(y_{t-w..t}) - \tau_{x,x'}))$$

$$f^{\geq}_{x,x'}(X_t, y_{t-w..t}) = sigmoid(\gamma_{x,x'}(\tau_{x,x'} - g(y_{t-w..t})))$$

- Normalized observations

$$f_{x,x'}(X_t, y_{t-w..t}) = \sigma^{-1}(g(y_{t-w..t}) - \mu)$$

Conditional Random Field (CRF)

- Feature extraction: estimate threshold and other param
  - E.g., logistic regression

$$\tau^*, \gamma^* = \text{argmax}_{\tau,\gamma} \Pr(x_{0..T} | y_{0..T}, \tau, \gamma)$$

- CRF Training: maximize conditional log-likelihood

$$\lambda^* = \text{argmax}_{\lambda} \Pr(x_{0..T} | y_{0..T}, \lambda)$$

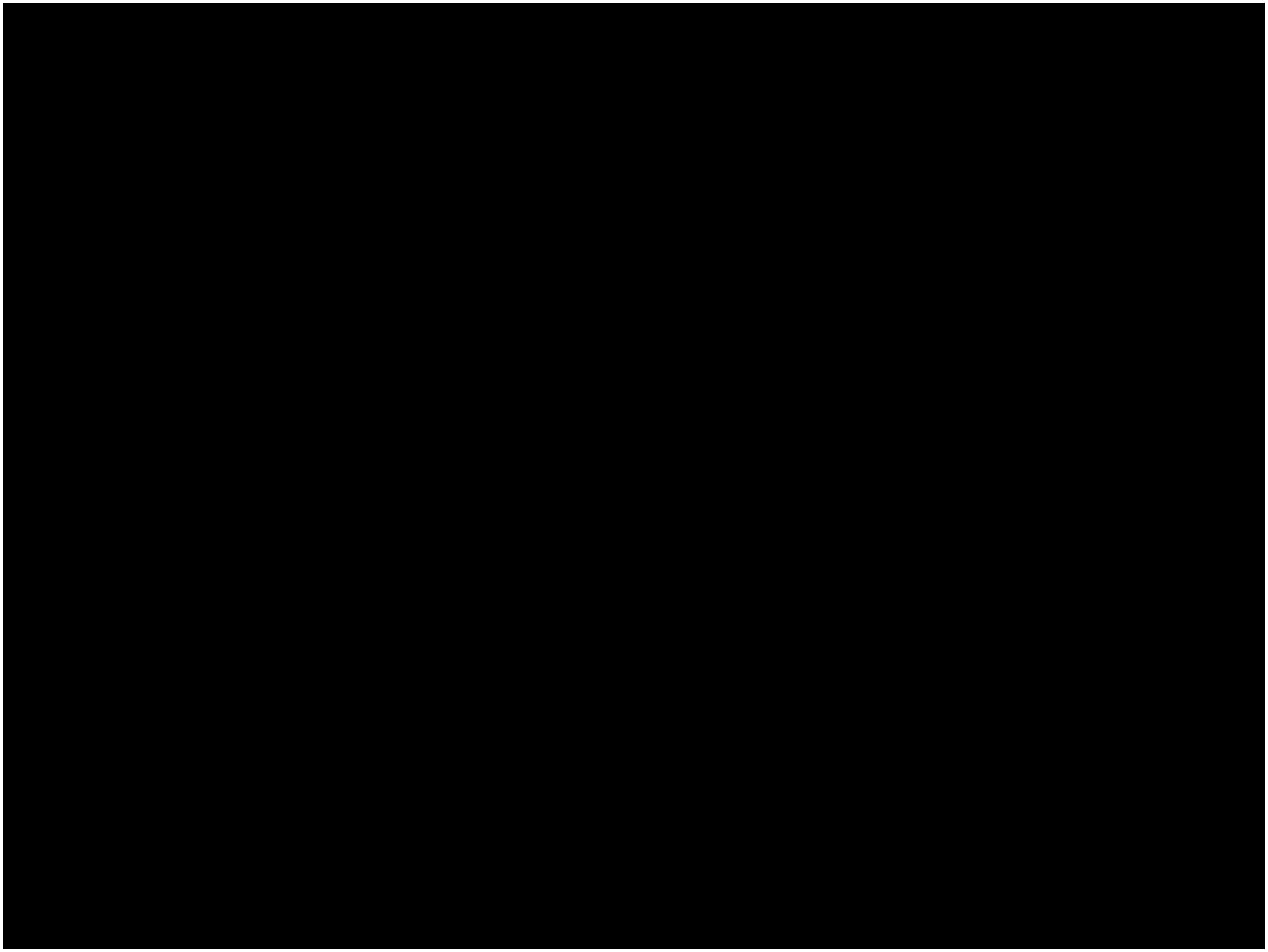
  - Concave optimization
  - Conjugate gradient

- Prediction:

$$x_t^* = \text{argmax}_{x} \Pr(X_t = x | y_{0..t}, \lambda^*)$$

  - Online and real-time

Demo



Evaluation Methodology

- Leave-one-out cross validation
- Performance measures:
  - Activity prediction

$$\text{accuracy} = \% \text{ of frames with correct activity label}$$

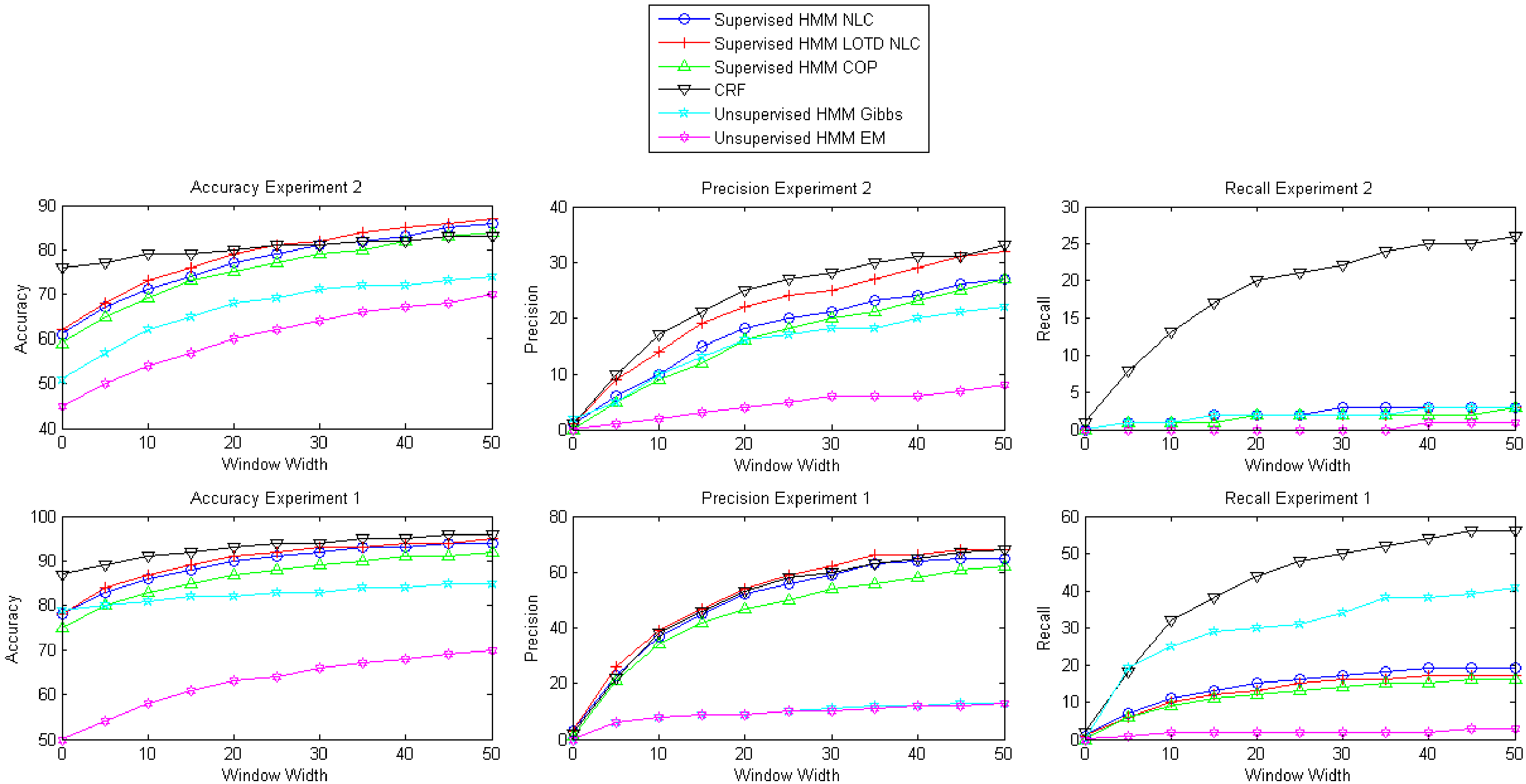
  - Activity changes:

$$\text{precision} = \frac{\# \text{ of correctly predicted changes}}{\# \text{ of predicted changes}}$$

$$\text{recall} = \frac{\# \text{ of correctly predicted changes}}{\# \text{ of changes}}$$

- A prediction is counted as correct if the corresponding activity (change) occurs within a window of $\pm$w
  - w $\in$ [0, 50]

Results



Supervised HMM NLC
Supervised HMM LOTD NLC
Supervised HMM COP
CRF
Unsupervised HMM Gibbs
Unsupervised HMM EM
Accuracy Experiment 2
Precision Experiment 2
Recall Experiment 2
Accuracy Experiment 1
Precision Experiment 1
Recall Experiment 1
Accuracy
Precision
Recall
Window Width

Results

- Accuracy in %, window size: 25

Model	NTW	ST	WF	TL	TR	WB	TRS	overall
HMM (ML)	65	88	95	96	92	95	85	91
CRF	96	82	98	89	80	94	70	93
HMM (EM)	100	14	48	94	90	97	0	61
HMM (Gibbs)	100	72	95	66	72	44	17	83

Model	NTW	ST	WF	TL	TR	WB	RT	SW	UR	DR	UC	DC	total
HMM (ML)	50	71	73	81	73	21	52	99	86	86	94	85	81
CRF	78	94	95	51	20	0	5	99	31	23	67	13	81
HMM (EM)	73	33	54	63	63	7	30	99	40	23	15	91	62
HMM (Gibbs)	100	42	77	69	81	0	62	91	57	52	46	2	69

Discussion

- Experiment 1: better results than Experiment 2
  - Participants perform the course twice
  - Less complex activities

- CRF: better results than HMM
  - Discriminative approach well suited
  - Sensor measurements violate conditional independence

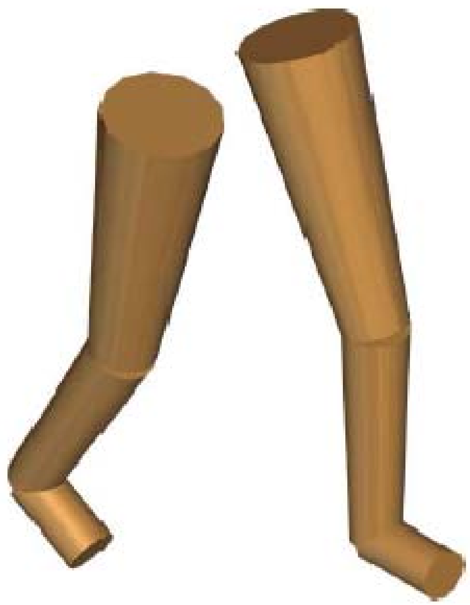
- Latent states obtained by unsupervised techniques:
  - Manually match each latent state with the activity it is most frequently predicted as
  - Sub-dividing or merging activities may improve the results

Outline

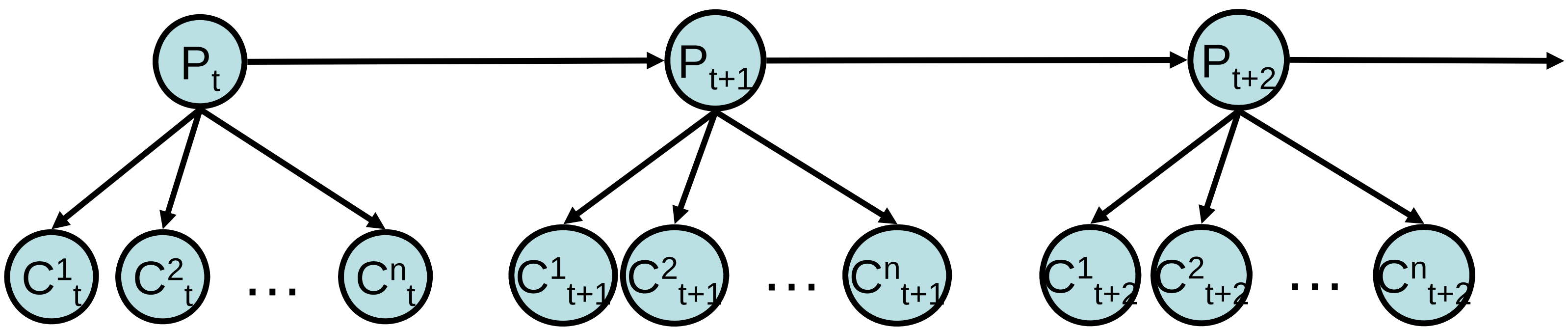
- Motivation
- Activity recognition
  - Experimental data
  - Algorithms (supervised/unsupervised HMMs and CRFs)
  - Results
- 3D lower limb tracking
  - Challenges
  - Trackers (RGB binocular and structured light cameras)
  - Results
- Conclusion & future work

Limb Tracking

- Estimate 3D pose
- Tapered cylinders defined by 21 parameters



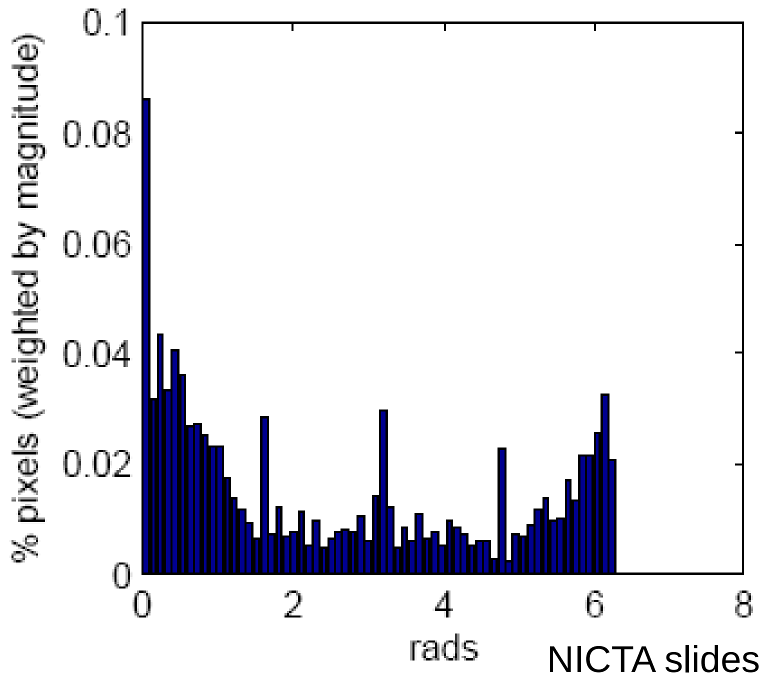
- HMM: particle filtering
  - infer pose P from cues C^i



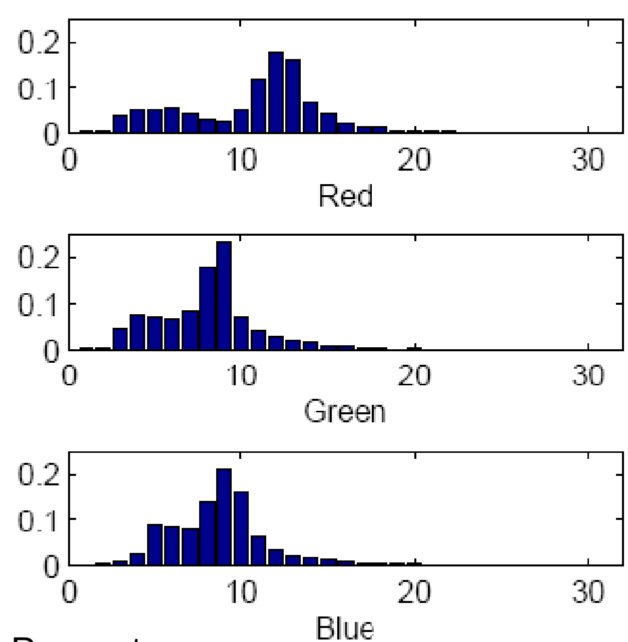
Appearance Model (Cues)

- For each limb segment:
  - Texture: histogram of oriented gradients (HOG)
  - Colour: histogram of colours (RGB)

Ex: histogram of oriented gradients



Ex: histogram of colours



Pose estimation

- HMM
  - Transition model:
    - Linear Gaussian (constant velocity)

$$\Pr(P_t | p_{t-1}, p_{t-2}) = N(2p_{t-1} - p_{t-2}, \Sigma)$$

  - Observation model:
    - Exponential distribution (difference of histograms)

$$\Pr(C_t^{1:n} | p_{1:t-1}) \propto exp(\textstyle\sum_{i=1}^{n} d_t^i)$$

$$\text{where} \quad d_t^i = |HoG_t^i - HoG_*^i|$$

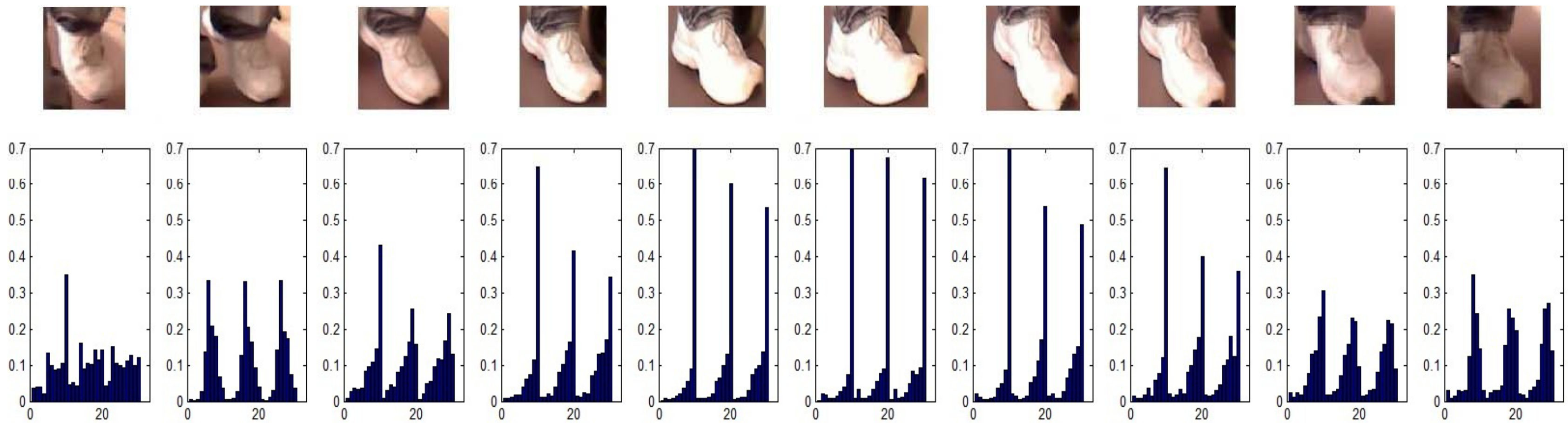
$$d_t^i = \textstyle\sum_{c \in \{R,G,B\}} |HoC_t^{i,c} - HoC_*^{i,c}|$$

- Pose estimation (Bayes' rule):

$$\Pr(P_t | C_{1:t}) \propto \int_{P_{t-1}} \Pr(P_{t-1} | C_{1:t-1}) \Pr(P_t | P_{t-1}) \Pr(C_t | P_t)$$

Challenges

- Changing lighting conditions



- Motion capture:

Sagittal plane

easy

Coronal plane

hard

Stereo Vision

- Challenges
  - Clothes with uniform texture → few features
  - Can't use window-based stereo matching

- Binocular RBG cameras
  - Treat each camera as a set of cues
  - Implicitly recover depth information by inference

- Structured light camera (Kinect)
  - Stereo matching based on cloud of infrared dots
  - Image: 2D array of depth measurements

Binocular RBG tracker

Good

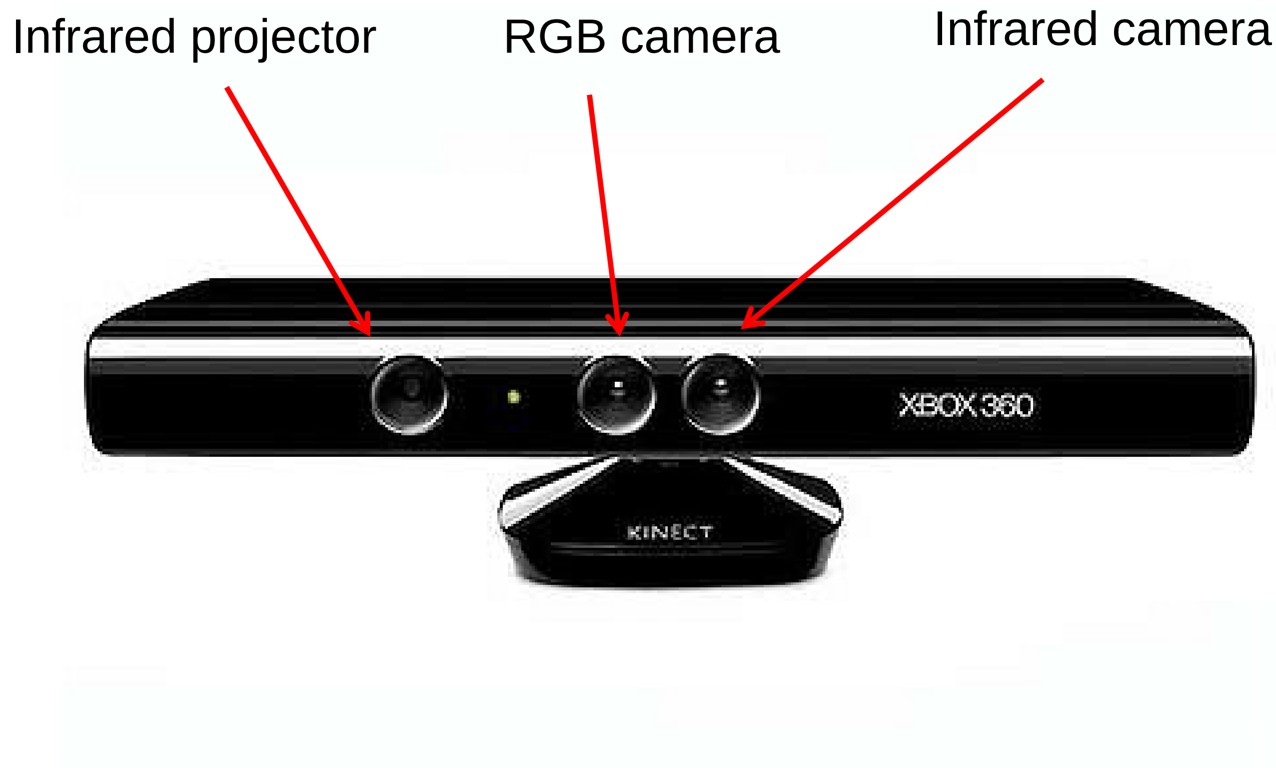
Not so good

Results

- Mean ankle error (cm)

Walk	Coord	1 Cam	2 Cams	2 Cams + color norm	Human
1	X	4.28	3.35	2.56	1.73
	Y	20.37	13.29	5.53	3.30
	Z	26.19	17.06	11.57	10.47
2	X	6.04	8.07	3.60	1.89
	Y	23.43	16.76	12.48	1.80
	Z	26.71	25.12	16.62	11.23
3	X	5.16	3.68	4.75	2.87
	Y	13.42	10.30	12.87	2.32
	Z	16.32	12.76	15.57	12.63

Kinect


Infrared projector
RGB camera
Infrared camera
XBOX 360
KINECT

Depth map output

Infrared image

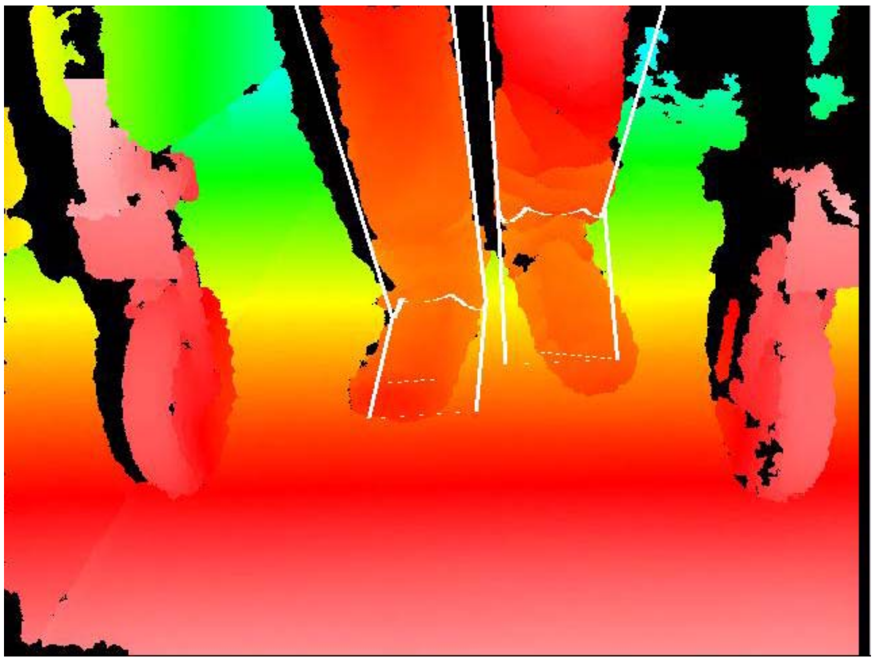
Gray scale depth map

Two cues

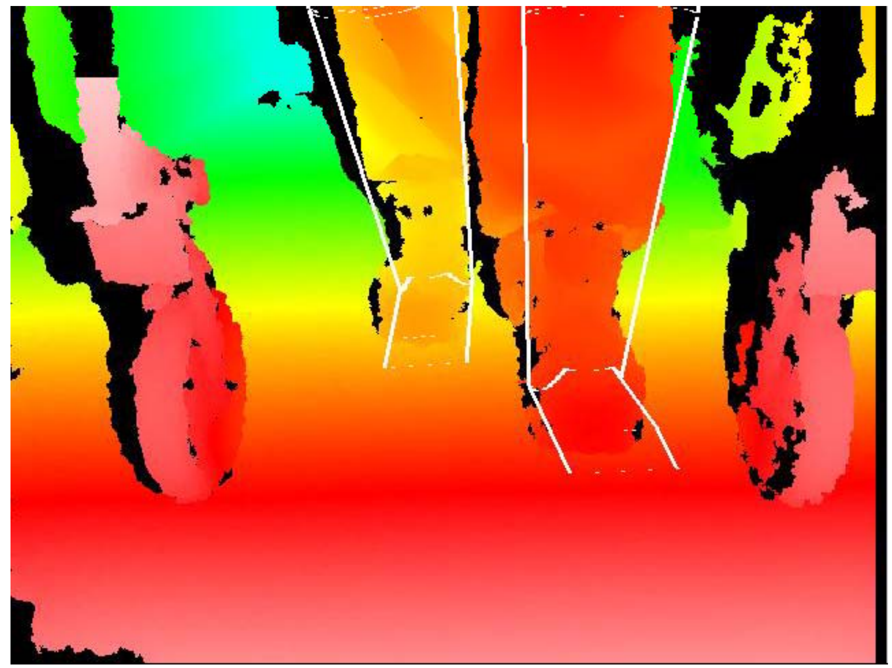
- 3D cue:
  - Compute 3D surface from Kinect depth map
  - d_1 = || cylinders - 3D surface ||
- Projection cue:
  - Compute cylinder depth map by projection in 2D
  - d_2 = || projected cylinders - Kinect depth map ||

Kinect-based tracker

Subject 1



Subject 2



Results

- Step length error (cm)

Subject	3D cue	Projection cue	Combined cues	RGB
1	4.67 ± 1.66	7.11 ± 3.11	2.73 ± 2.31	18.80 ± 14.21
2	3.88 ± 4.69	15.46 ± 6.98	3.87 ± 2.27	22.43 ± 13.93
3	4.60 ± 2.28	4.77 ± 4.23	3.48 ± 2.02	20.12 ± 12.15

- Step width error (cm)

Subject	3D cue	Projection cue	Combined cues	RGB
1	8.71 ± 1.07	2.37 ± 1.91	2.87 ± 2.50	7.26 ± 6.83
2	7.00 ± 2.01	3.45 ± 2.43	3.04 ± 2.52	11.97 ± 8.51
3	4.56 ± 2.48	2.36 ± 1.22	1.75 ± 1.17	12.23 ± 7.54

Conclusion

- Automated Activity Recognition
  - Good accuracy
  - Experiments with walker users
  - Best approach: CRF with sigmoid features

- Lower limb tracking
  - Binocular RGB system: implicit depth, but subject to lighting conditions
  - Structured light camera (Kinect): quite promising