Enhanced Protection of Vulnerable Road Users

Miguel Ángel Sotelo
President. IEEE ITS Society

miguel.sotelo@uah.es
Full Professor
University of Alcalá (UAH)
SPAIN

Content

- **Motivation**
- Proposed approach
- VRU pose measurement
- Dimensionality reduction and Prediction
- Experimental results
- Conclusions and Future Work

Motivation

- **Getting to understand the intent of an observed agent (VRU):**

It is of paramount interest in a large variety of domains that involve some sort of collaborative and competitive scenarios

Robotics

Surveillance

HMI

Gaming

Intelligent Vehicles

Motivation

- **VRU Path Prediction in the Automotive:**
  - Further improvement in state-of-the-art ADAS by means of action classification
    - Walking, St[illegible] [illegible]ding-in
  - I[illegible] [illegible]ffective
    interven[illegible]ns
  - Initiation of emergency braking 0.16 s in advance can potentially reduce severity of accidents injuries by 50%
  - Early recognition of pedestrians stopping actions can provide more accurate last-second active interventions

Strong gains are expected in the performance and reliability of active VRU protection systems

Motivation

IEEE SPECTRUM

The Big Problem With Self-Driving Cars Is People

And we'll go out of our way to make the problem worse

By **RODNEY BROOKS** Posted 27 Jul 2017 | 15:00 GMT

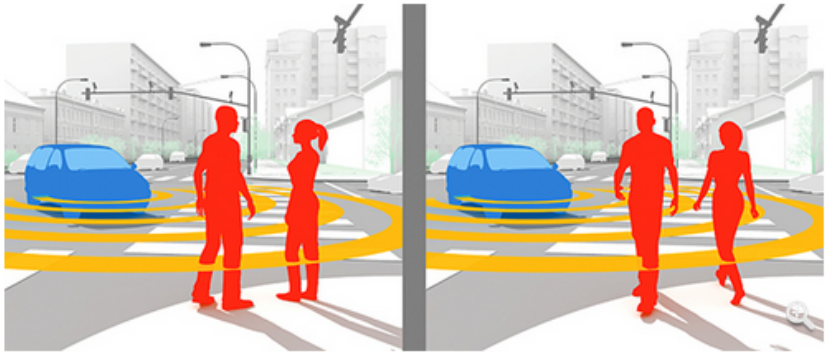


Illustration: Bryan Christie Design

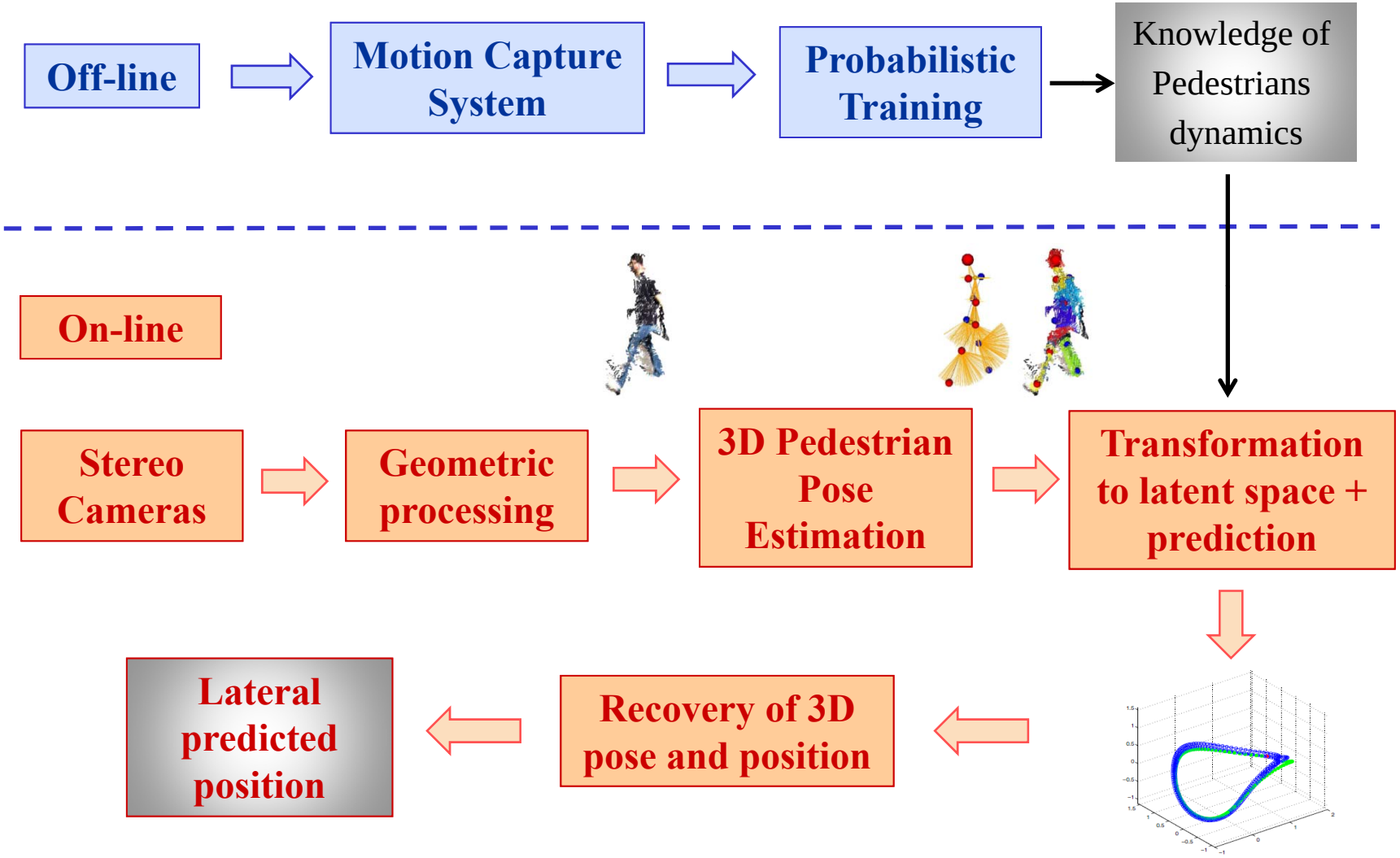
Reading Body Language: A purely interpretive problem that self-driving cars cannot yet solve is that of making sense of the way people hold themselves and move. For instance, a self-driving car could not tell what any human driver could take in at a glance: The couple conversing animatedly near the curb [left] are not about to wander into traffic. If, however, one person turns away from the other and in the direction of the street, it means she's about to cross [right].

Content

- Motivation
- **Proposed approach**
- Pedestrian pose measurement
- Dimensionality reduction and Prediction
- Experimental results
- Conclusions and Future Work

Proposed Approach

Global Scheme



Content

- Motivation
- Proposed approach
- **Pedestrian pose measurement**
- Dimensionality reduction and Prediction
- Experimental results
- Conclusions and Future Work

Pedestrian Pose Measurement

Goal

Estimation of a 3D pedestrian skeleton containing relevant body joints providing augmented features for path prediction

Steps

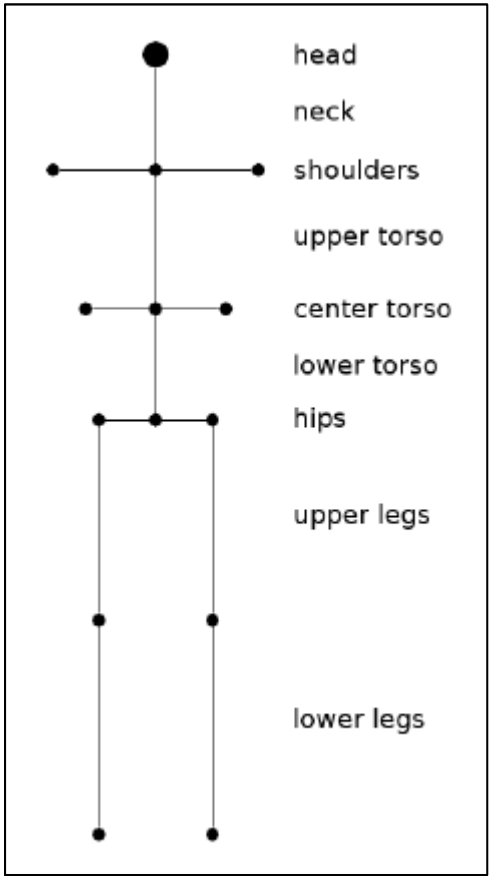
1. Computation of point cloud
2. Pedestrian extraction
3. Computation of 3D body joints (skeleton)

Inputs

Stereovision sequences obtained from different data-sets containing walking and stopping behaviors

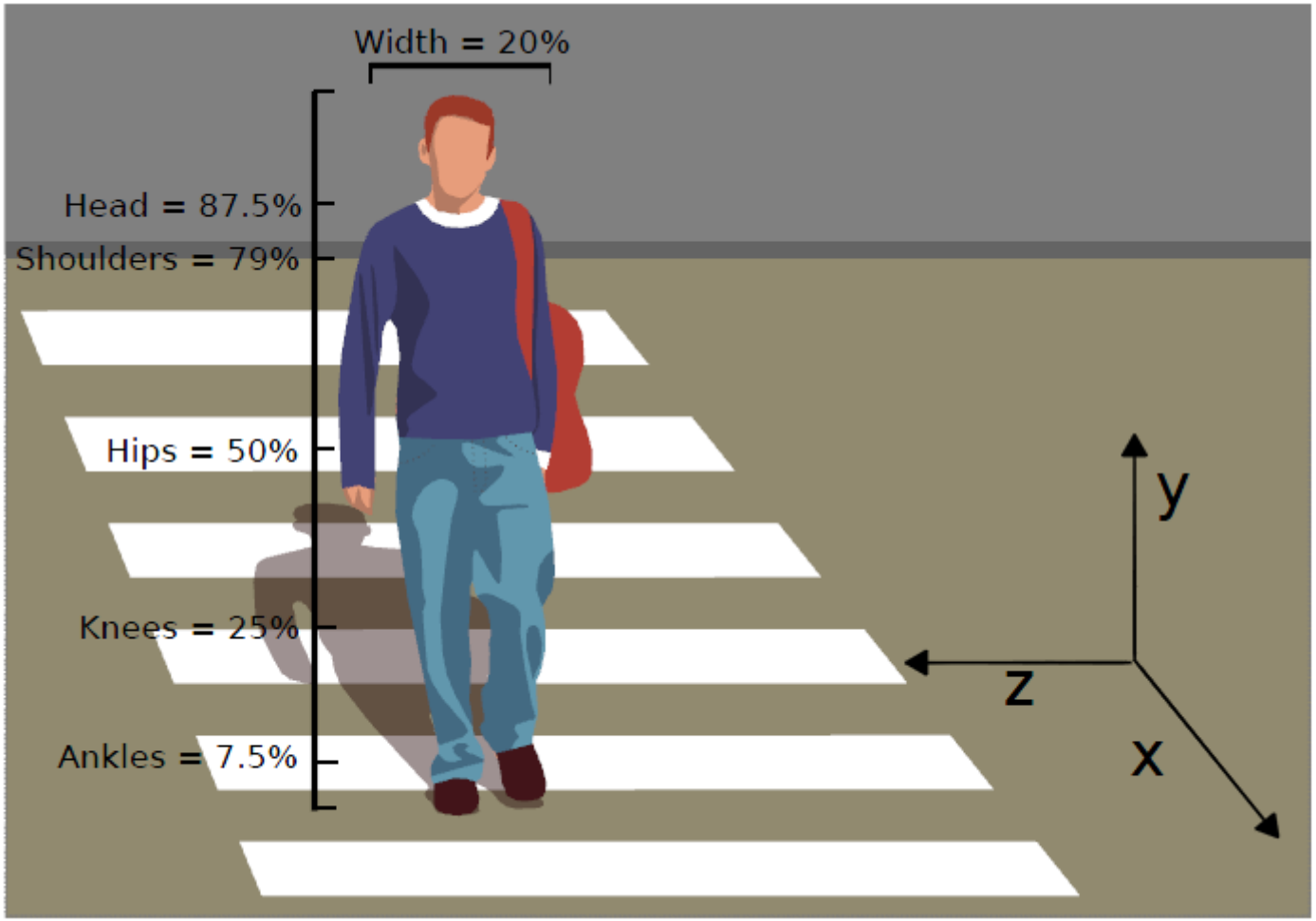
Pedestrian Pose Measurement

Pedestrian Skeleton considered in this research



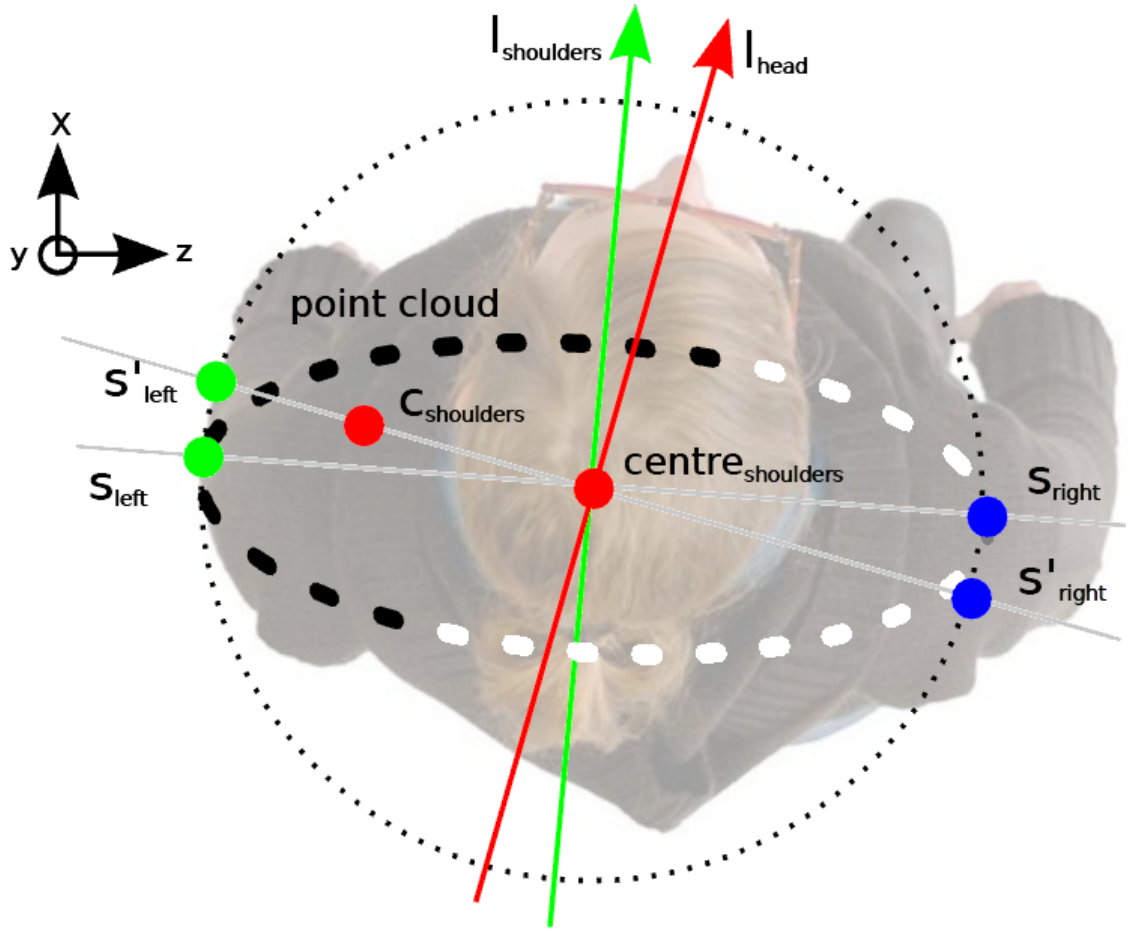
Pedestrian Pose Measurement

Pedestrian Skeleton – Anthropometric Proportions



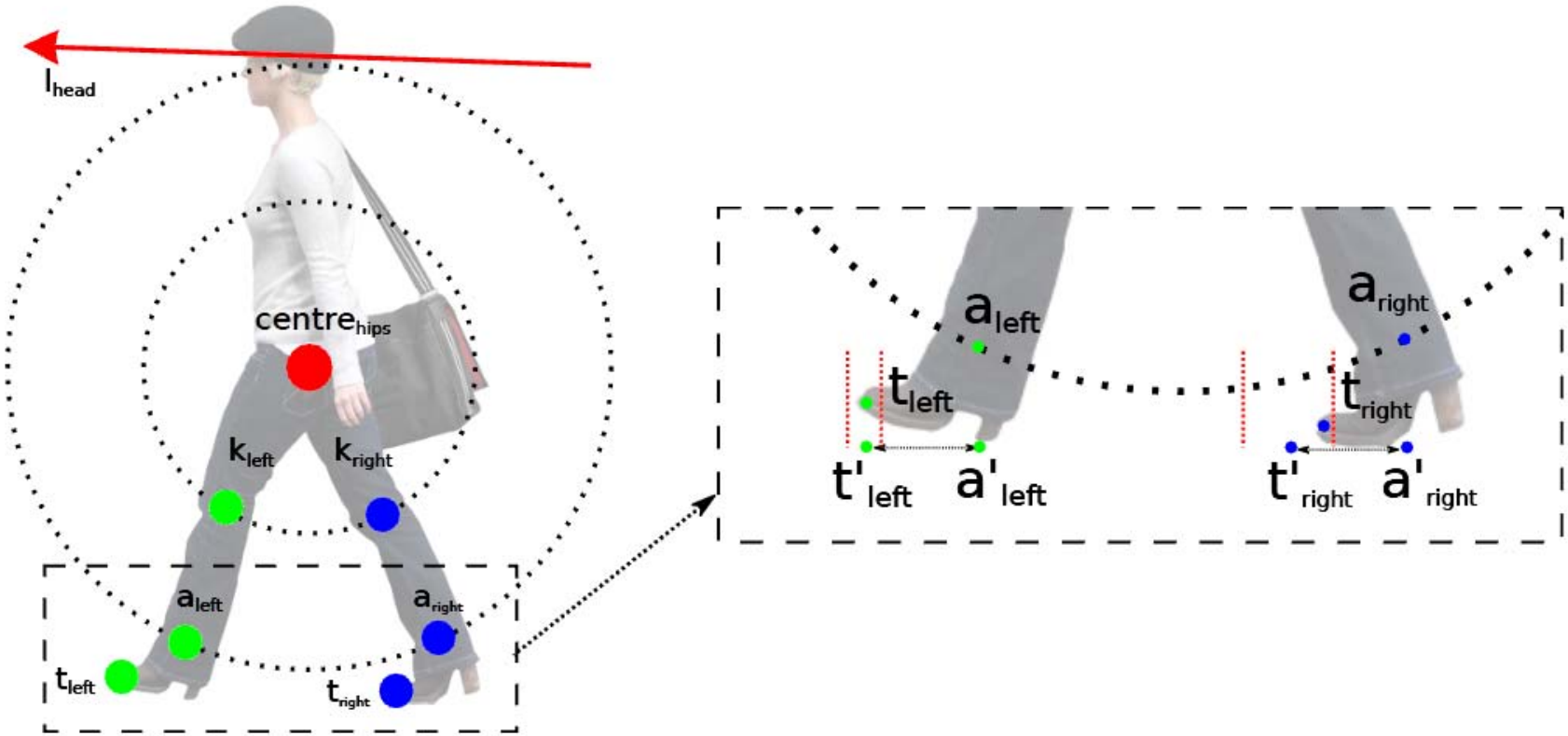
Pedestrian Pose Measurement

Pedestrian Shoulders Estimation



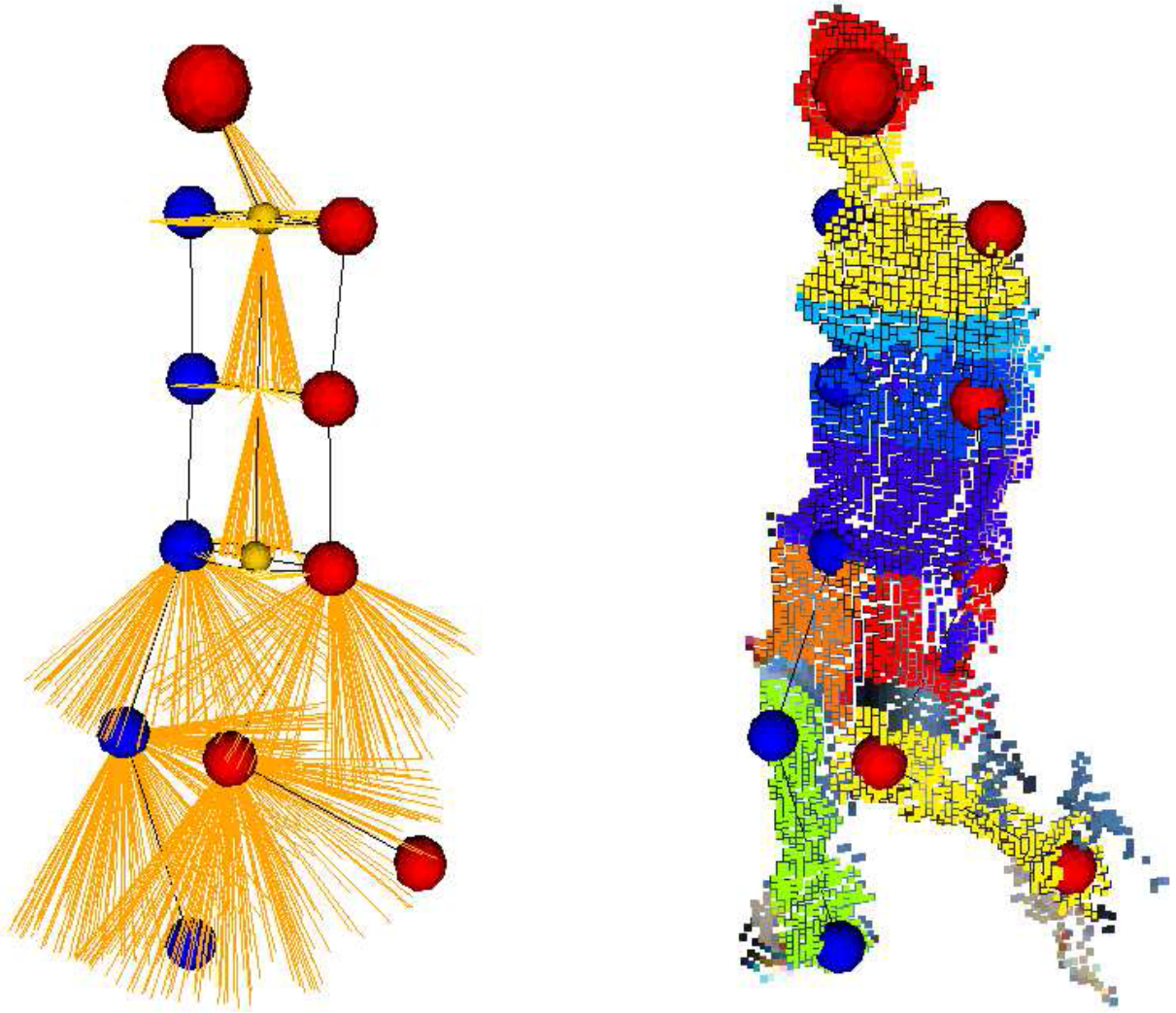
Pedestrian Pose Measurement

Pedestrian Lower Limbs Estimation



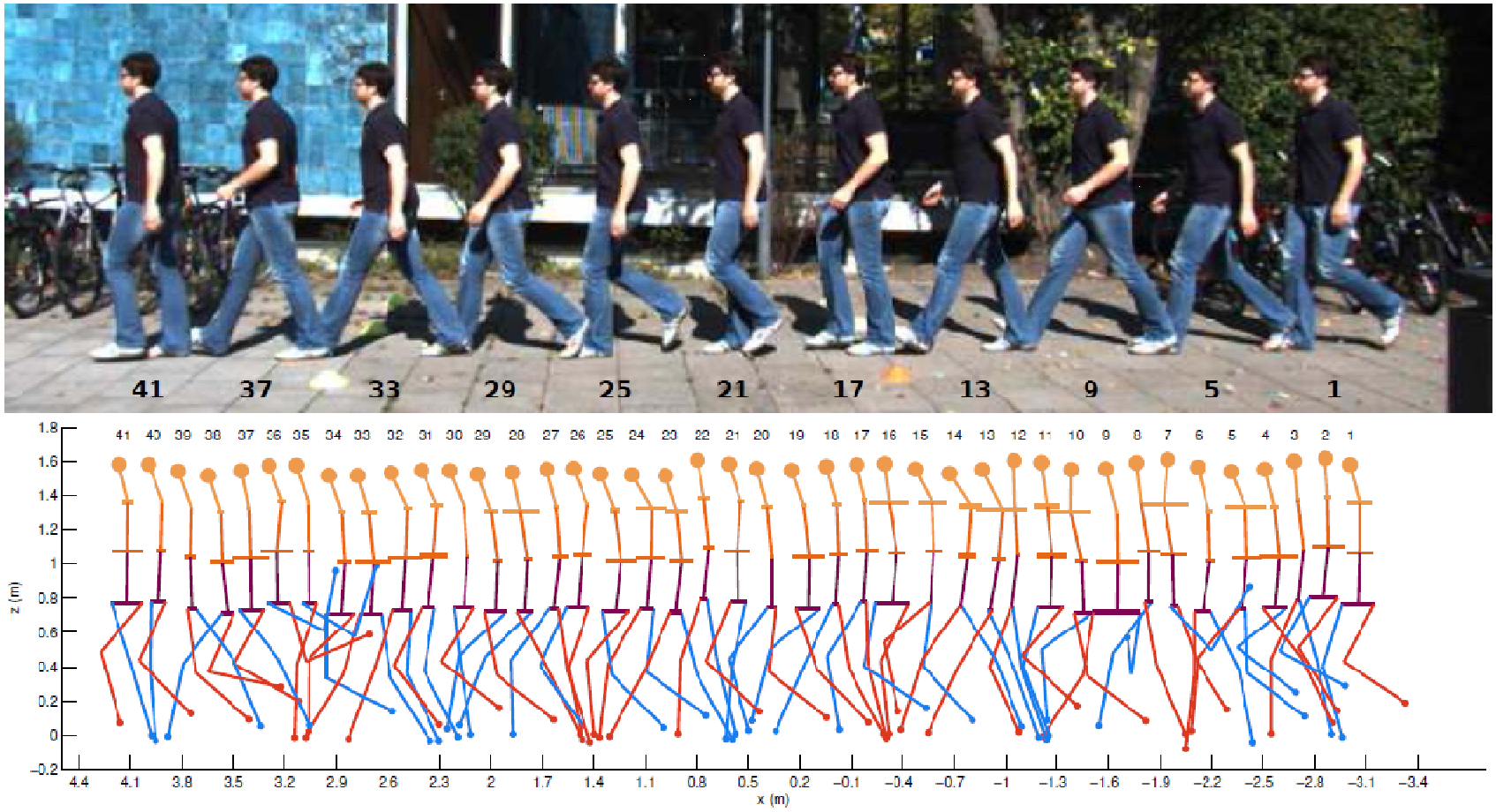
Pedestrian Pose Measurement

Method for Joints Extraction - Example



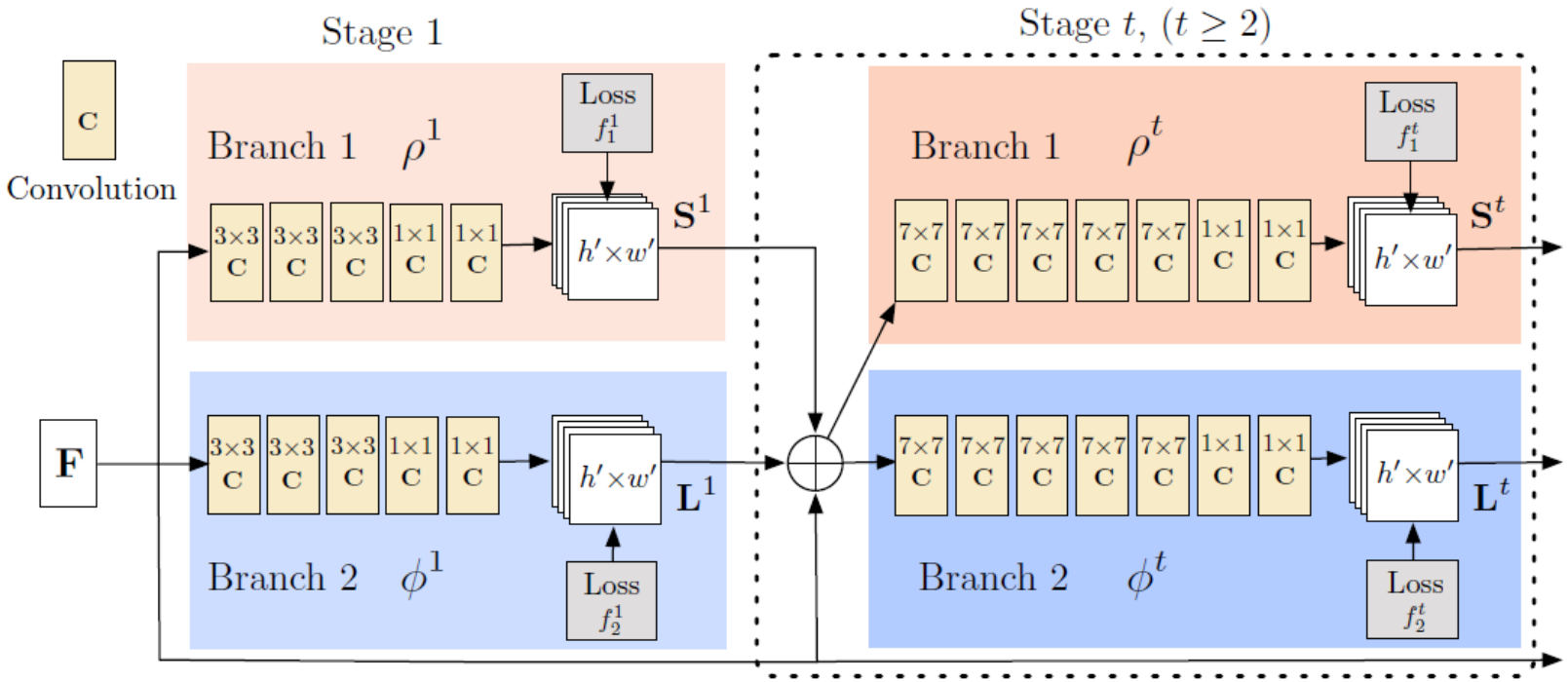
Pedestrian Pose Measurement

Method for Joints Extraction – Results



Pedestrian Pose Measurement

Body parts detection using Deep Learning

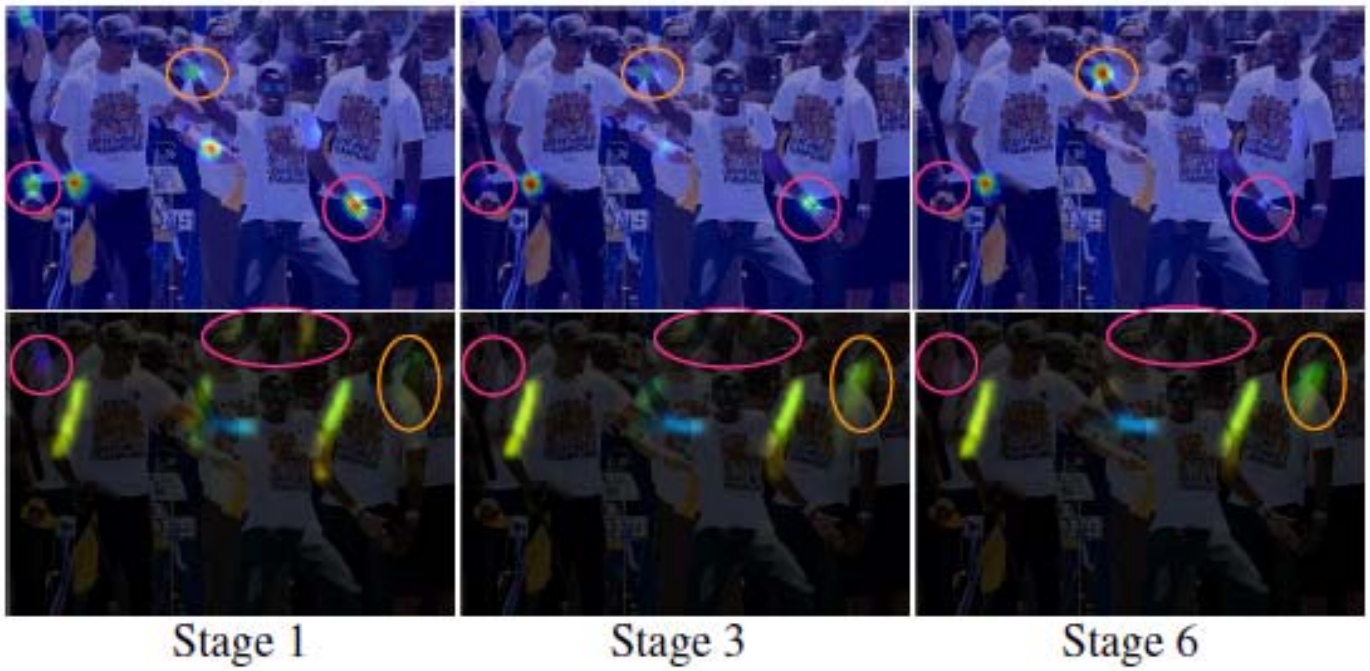


St: Confidence Maps

Lt: Parts Affinity Maps (PAF) (Z. Cao. CVPR 2017)

Pedestrian Pose Measurement

Body parts detection using Deep Learning



Confidence maps of the right wrist (first row) and PAFs (second row) of right forearm across stages

Pedestrian Pose Measurement

Body parts detection using Deep Learning

Pedestrian Pose Measurement

Body parts detection using Deep Learning

Content

- Motivation
- Proposed approach
- Pedestrian pose measurement
- **Dimensionality reduction and Prediction**
- Experimental results
- Conclusions and Future Work

Dimensionality Reduction

Major Goal

Learning the pedestrians motion dynamics by reducing the dimensionality of input data and to make it more manageable and interpretable in a low dimensional latent embedding

Method considered

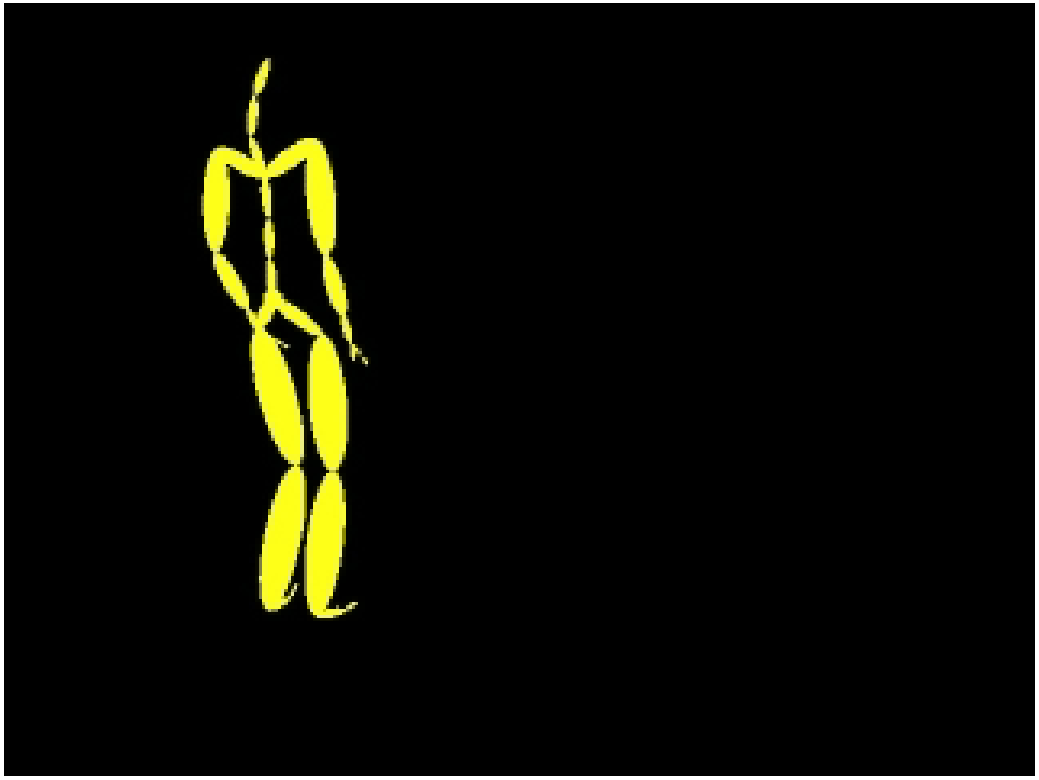
GPDM (Gaussian Process with Dynamical Model)

Input data-set

CMU (41 body joints obtained with motion capture syst.)

Dimensionality Reduction

CMU Sequence Example



Dimensionality Reduction

GPDM (1/3)

1. Nonlinear probabilistic generalization of PCA: marginalization of the mapping (W) and optimization of the latent variables (X)
2. Conditional density in GPDM for a centered observed data Y, latent variable X, and kernel hyperparameters θ

$$p(Y|X,\theta,W) = \frac{|W|^N}{\sqrt{(2\pi)^{ND}|K_Y|^D}} exp(-\frac{1}{2}tr(K_Y^{-1}YW^2Y^T))$$

where W is a scaling matrix and K is a kernel matrix with:

$$k(x_i,x_j) = \theta_1 exp(\frac{-\theta_2}{2}(x_i - x_j)^T(x_i - x_j)) + \theta_3\delta_{i,j}$$

where $\delta_{i,j}$ is the Kronecker delta function

Dimensionality Reduction

GPDM (2/3)

3. Dynamic mapping from latent coordinates:

$$p(X|\beta) = \frac{p(x_1)}{\sqrt{(2\pi)^{(N-1)d}|K_X|^d}} exp(-\frac{1}{2}tr(K_X^{-1}X_{out}X_{out}^T))$$

where $X_{out}=[X_2, \ldots, X_N]$, d is the model dimension and K_X is the kernel built from $\{X_1, \ldots, X_{N-1}\}$ with:

$$k(x_i,x_j) = \beta_1 exp(\frac{-\beta_2}{2}(x_i - x_j)^T(x_i - x_j)) + \beta_3 x_i^T x_j + \beta_4 \delta_{i,j}$$

where β_i are hyperparameters

4. Minimization of $-ln\ p(X,\theta,\beta,W|Y)$ using SCG

$$\mathscr{L} = \mathscr{L}_Y + \mathscr{L}_X + \sum_j ln\theta_j + \frac{1}{2\kappa^2}tr(W^2) + \sum_j ln\beta_j$$

Dimensionality Reduction

GPDM (3/3)

where

$$\mathscr{L}_Y = \frac{D}{2} ln|K_Y| + \frac{1}{2} tr(K_Y^{-1} Y W^2 Y^T) - N ln|W|$$

$$\mathscr{L}_X = \frac{d}{2} ln|K_X| + \frac{1}{2} tr(K_X^{-1} X_{out} X_{out}^T) + \frac{1}{2} x_1^T x_1$$

Prediction

1. Predicted latent position in the next frame (k_X is a column vector)

$$\mu_X(x) = X_{out}^T K_X^{-1} k_X(x)$$

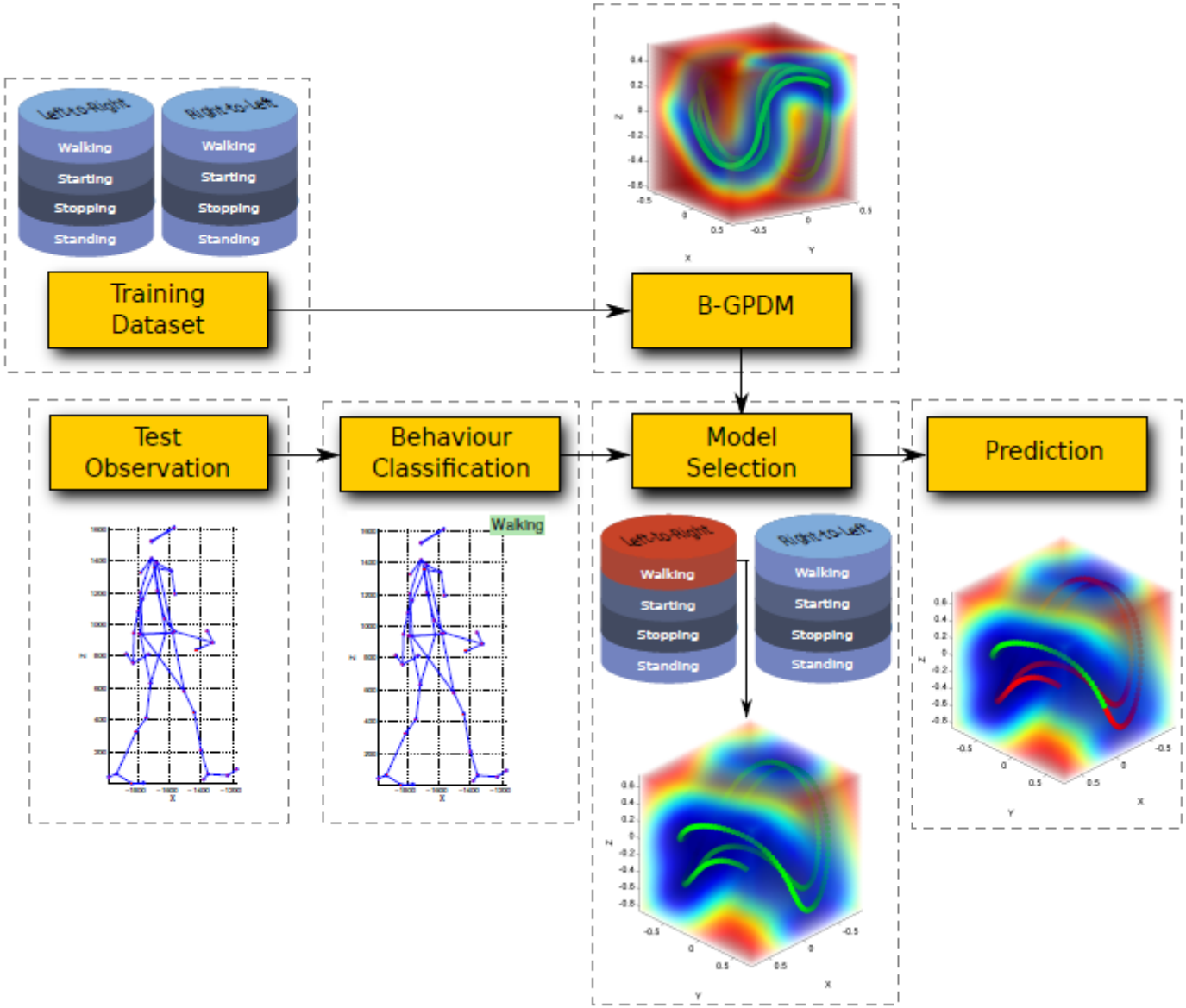
2. Reconstructed 3D pose (k_Y is a column vector)

$$\mu = Y^T K_Y^{-1} k_Y(x)$$

Content

- Motivation
- Proposed approach
- Pedestrian pose measurement
- Dimensionality reduction and Prediction
- **Experimental results**
- Conclusions and Future Work

General Method - Overview



Datasets

Frame Rate of training data

CMU data-set: 120 fps (frames per second)

Feature Vector

UAH: 11 body joints (3D pose + 3D motion: 11x6)

CMU: 41 (we consider only the 11 joints in common with UAH)

Pedestrian Behaviors

Walking (toward the curb to enter the road lane)

Stopping (at the curb)

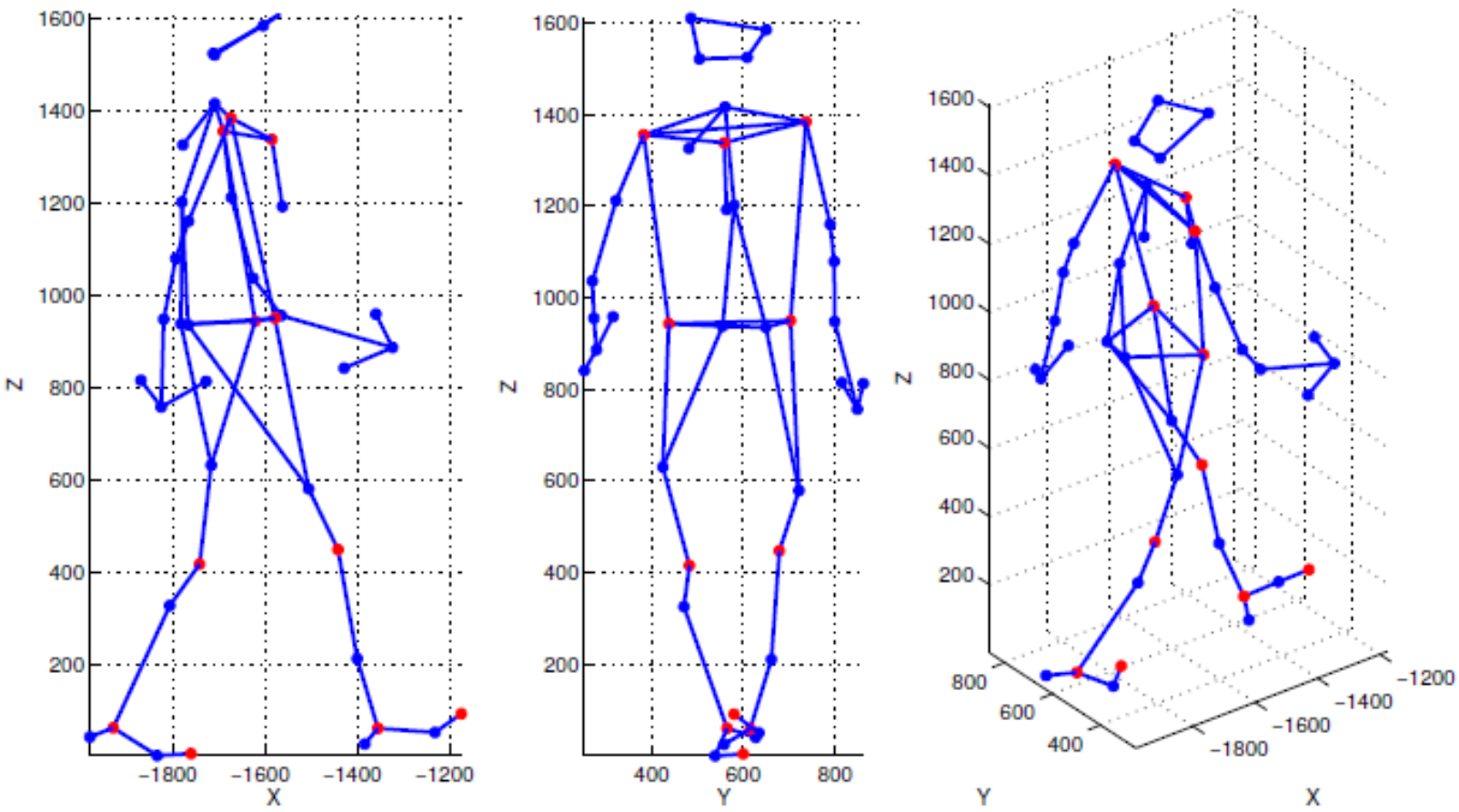
Starting (from the curb)

Standing

Prediction Horizon

Up to 1.0 s

Datasets



Blue markers (41 joints):	**CMU dataset**
Red markers (11 joints):	**UAH dataset**

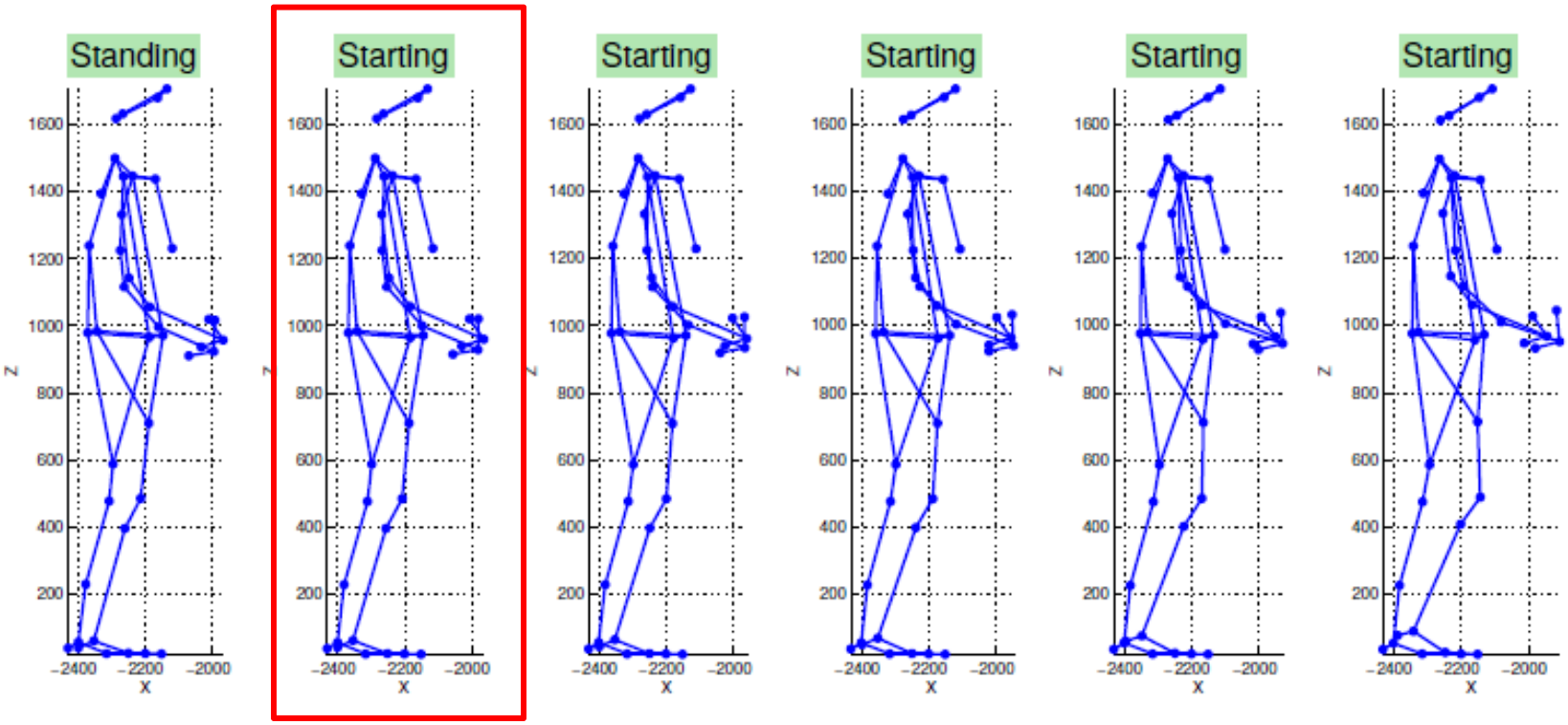
Experimental Results

UAH Dataset - Breakdown

	Orientation	Walking	Starting	Stopping	Standing	Total
Sequences	Left-to-right	240	142	56	224	662
Sequences	Right-to-left	191	121	27	156	495
Total sequences		431	263	83	380	1157
Pedestrian poses	Left-to-right	107324	10732	2522	43151	163729
Pedestrian poses	Right-to-left	95113	10940	1276	31412	138741
Total pedestrian poses		202437	21672	3798	74563	302470

Manual Labelling of Pedestrian Poses

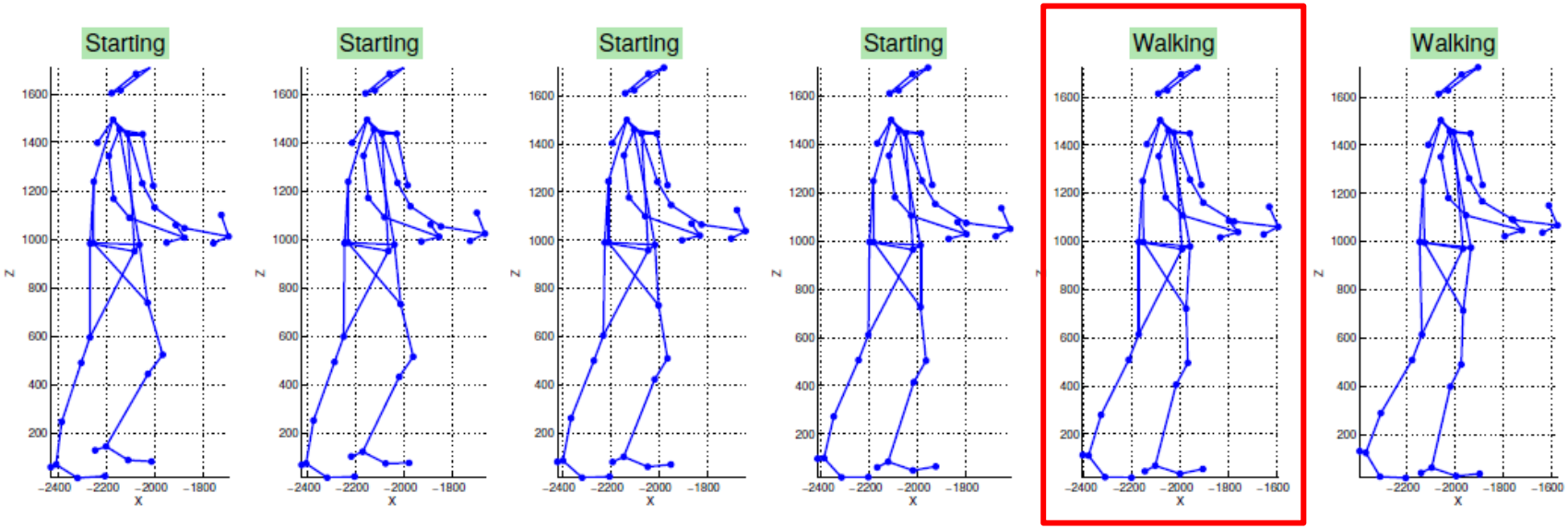
Standing-Starting Transition



Criteria: movement of leg forward

Manual Labelling of Pedestrian Poses

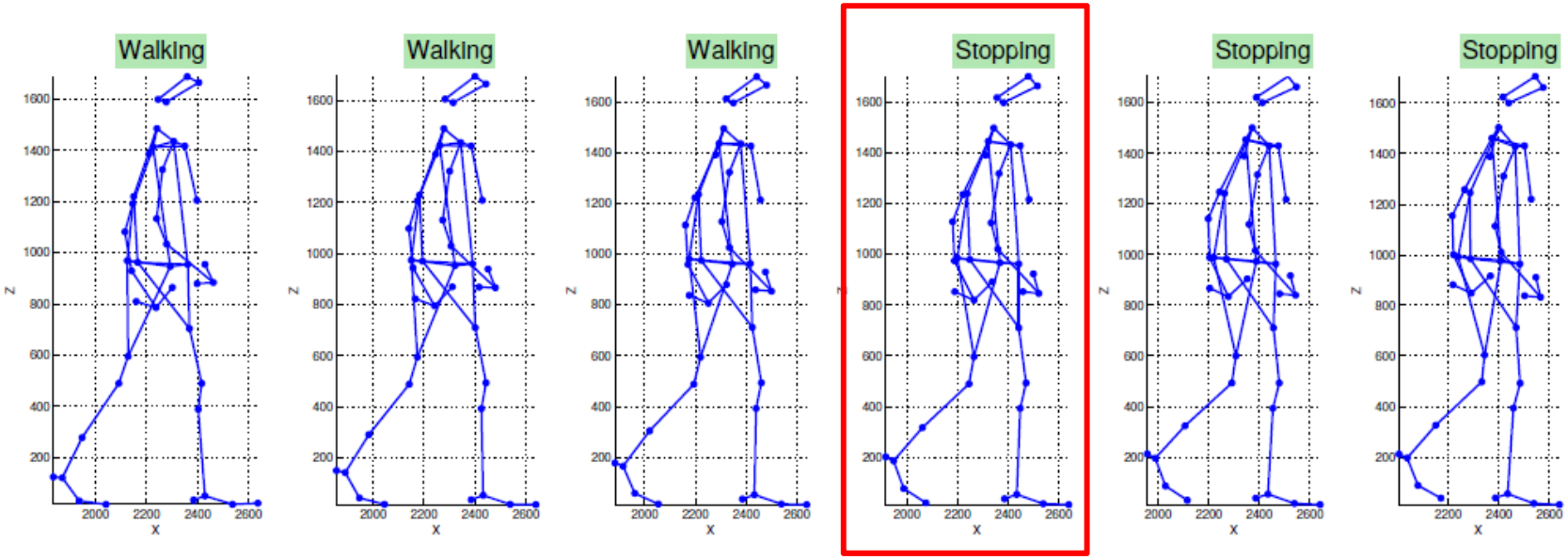
Starting-Walking Transition



Criteria: pedestrian places foot on the ground after one stride

Manual Labelling of Pedestrian Poses

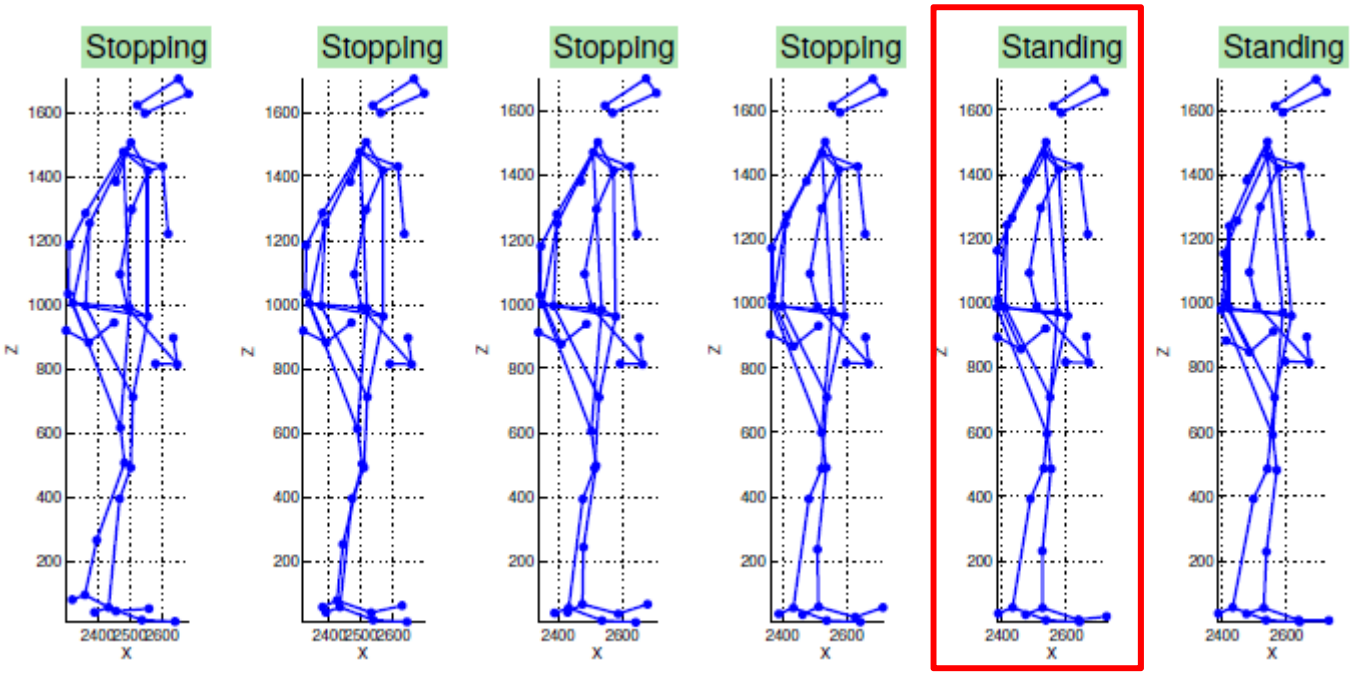
Walking-Stopping Transition



Criteria: pedestrian places foot on the ground in the final stride before full stop

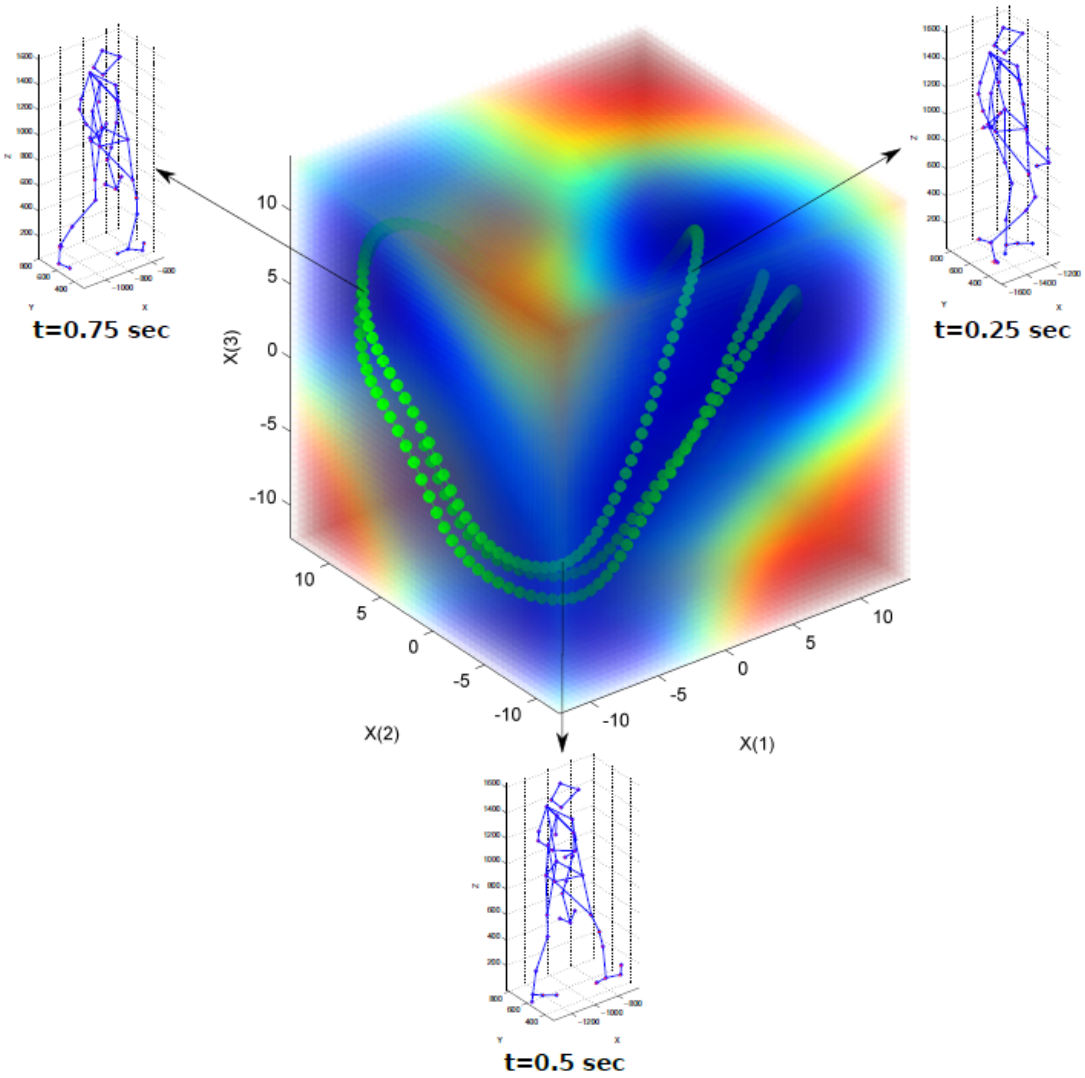
Manual Labelling of Pedestrian Poses

Stopping-Standing Transition



Criteria: pedestrian comes to a full stop

Learning Pedestrian Motion



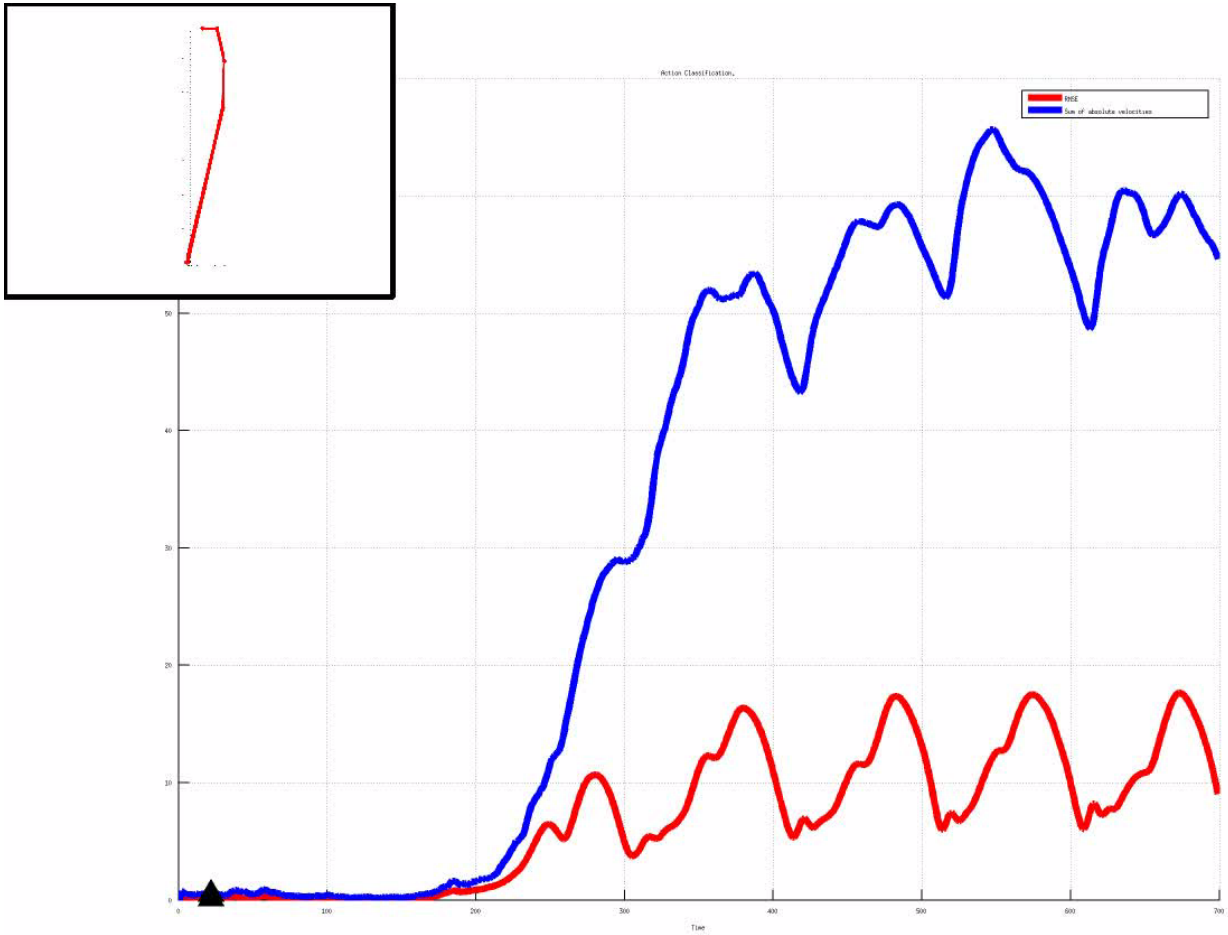
Pedestrian walking 6 steps (using B-GPDM)

Green: projection of pedestrian motion sequence onto the subspace.

Variance in color code: cold (small) to warm (big)

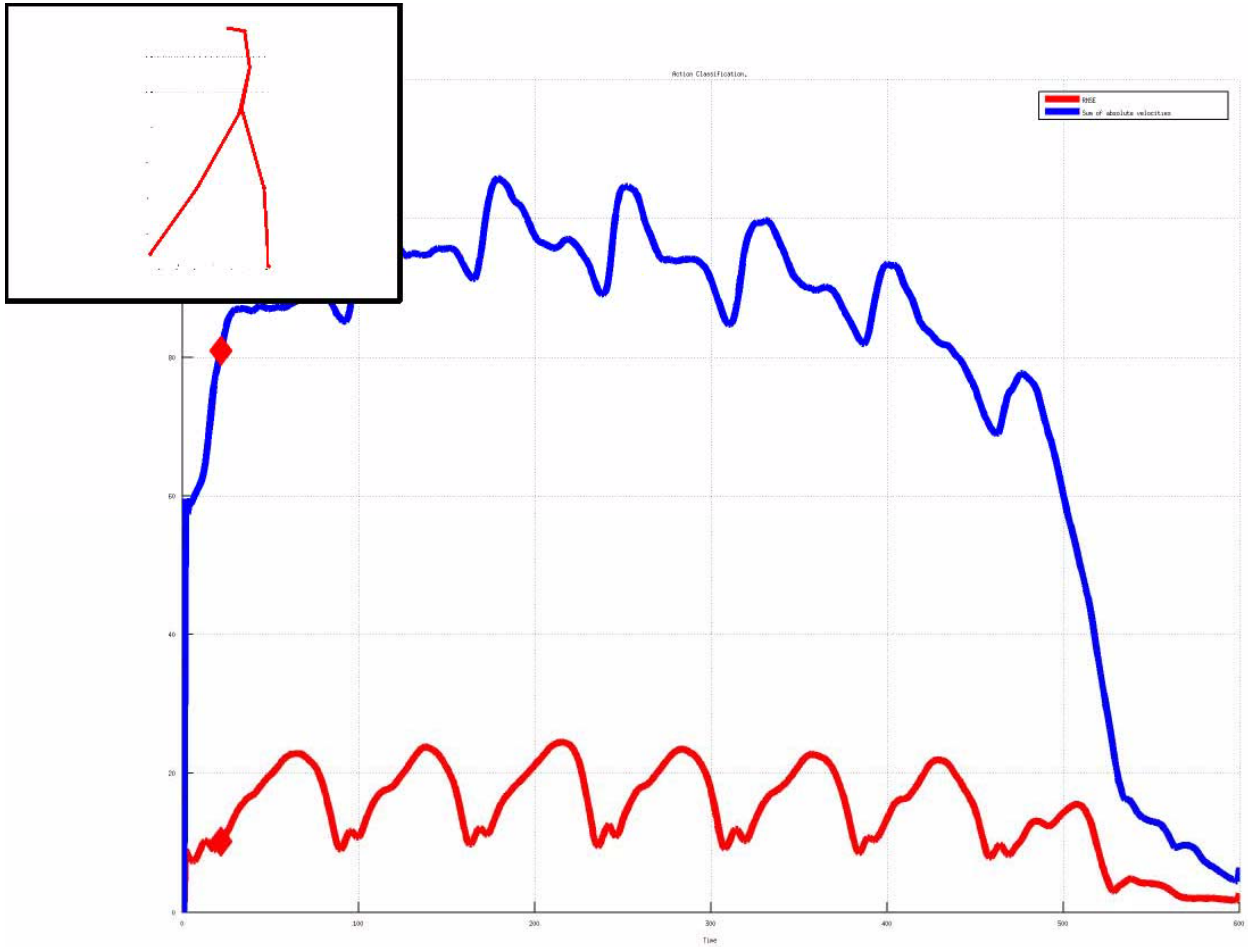
Activity Recognition

Clues for prediction of "Starting" behaviors



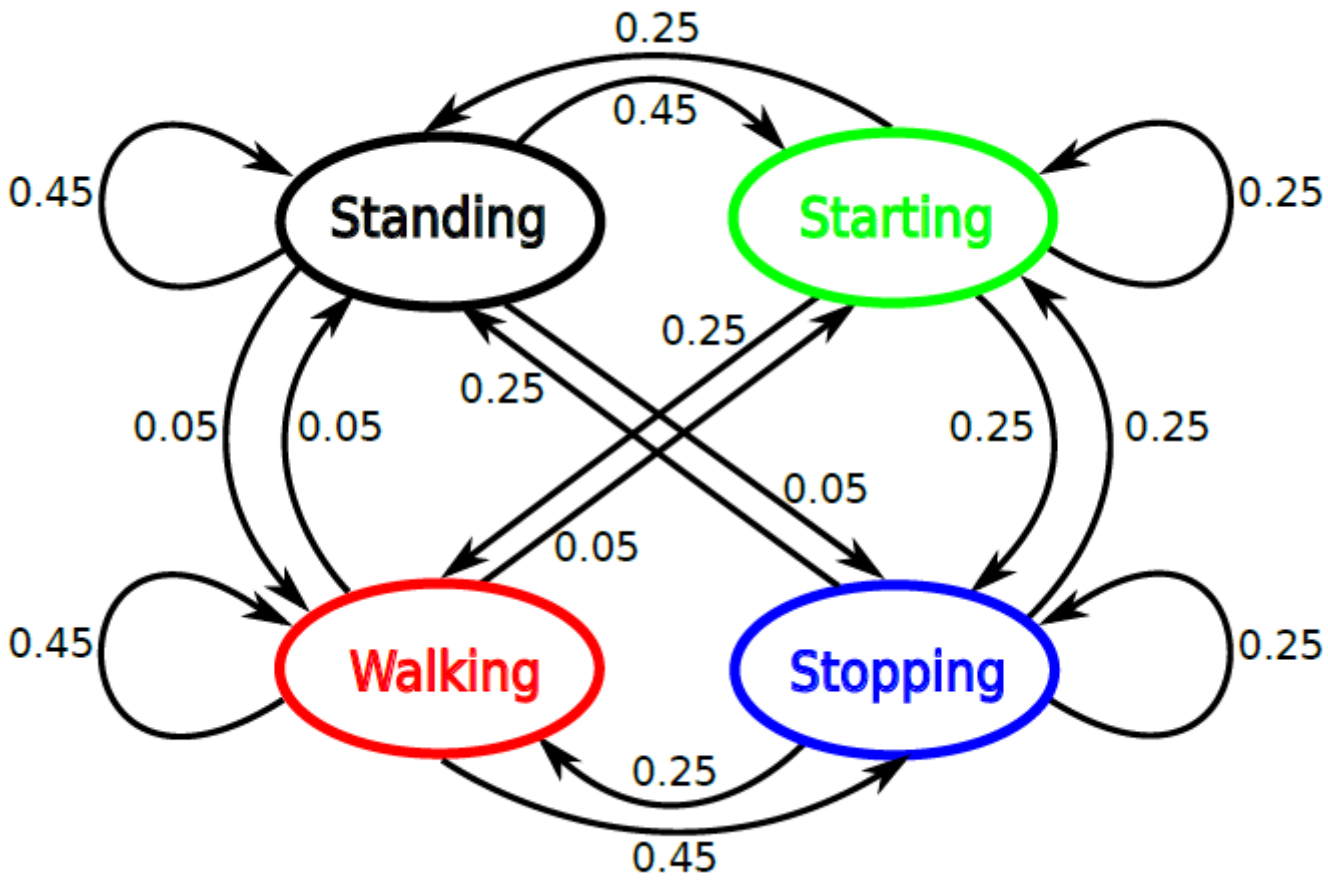
Activity Recognition

Clues for prediction of "Stopping" behaviors



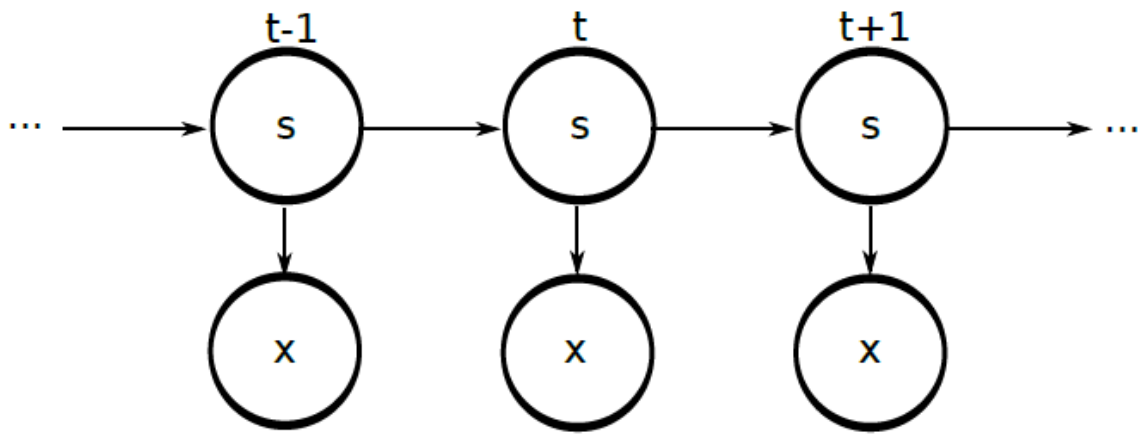
Activity Recognition

Pedestrian activity modelled as HMM



Activity Recognition

Pedestrian activity modelled as HMM



$$P(\mathbf{s}_j^t|\mathbf{x}^t) = \frac{P(\mathbf{x}^t|\mathbf{s}_j^t)P(\mathbf{s}_j^t)}{\sum_{i=1}^{4} P(\mathbf{x}^t|\mathbf{s}_i^t)P(\mathbf{s}_i^t)}$$

Activity Recognition

Pedestrian activity modelled as HMM

Prior

$$P(\mathbf{s}_j^t) \propto \max_{i=1}^{4}[P(\mathbf{s}_j^t|\mathbf{s}_i^{t-1})P(\mathbf{s}_i^{t-1}|\mathbf{x}^{t-1})], \quad t > 1$$

Emission (Likelihood)

$$P(\mathbf{x}^t|\mathbf{s}_j^t) \propto \max_{i=1}^{N}\left(\frac{1}{1+\alpha_i} + \frac{1}{1+\beta_i}\right)$$

α: SSE for pedestrian pose

β: SSE for pedestrian displacement

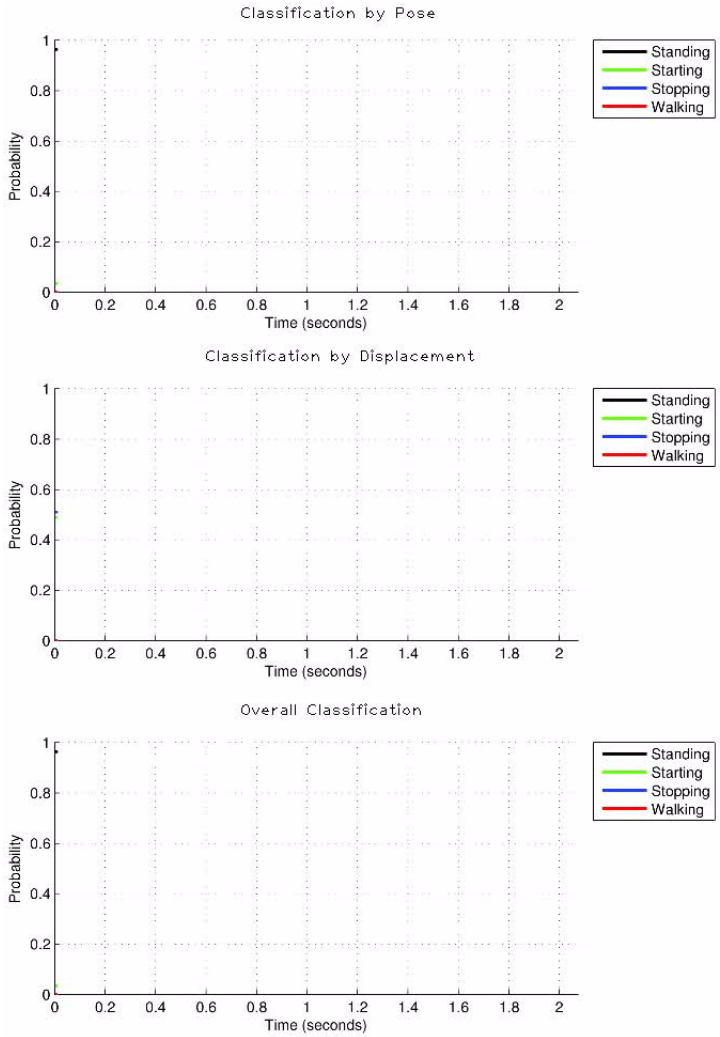
Activity Recognition

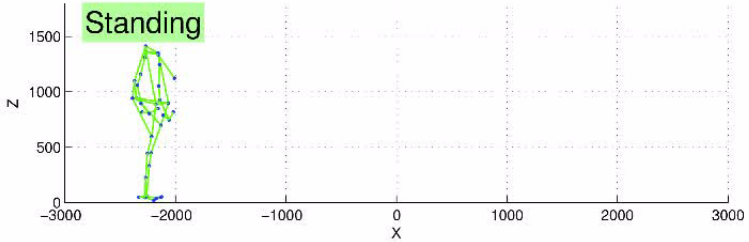
Recognition results for different features and #joints

Features		Pose + Disp		Pose		Disp	
Joints		41	11	41	11	41	11
Overall Accuracy		90.69%	**95.13%**	88.39%	91.28%	94.76%	94.23%
Precision	Standing	89.77%	97.51%	88.93%	95.54%	97.27%	**98.04%**
	Starting	77.88%	**87.57%**	66.30%	79.38%	82.79%	83.96%
	Stopping	44.59%	**53.51%**	41.78%	40.06%	35.79%	35.72%
	Walking	92.50%	**95.57%**	89.89%	91.35%	94.90%	94.81%
Recall	Standing	88.85%	96.97%	84.31%	87.86%	97.01%	**97.19%**
	Starting	48.60%	**61.49%**	31.33%	39.14%	52.62%	54.90%
	Stopping	36.45%	**51.38%**	32.87%	37.11%	41.14%	40.90%
	Walking	96.90%	98.88%	97.04%	**99.13%**	98.43%	98.36%
F1-Score	Standing	89.31%	97.24%	86.56%	91.54%	97.14%	**97.61%**
	Starting	59.85%	**72.25%**	42.55%	52.43%	64.34%	66.39%
	Stopping	40.11%	**52.42%**	36.79%	38.53%	38.28%	38.13%
	Walking	94.65%	**97.20%**	93.33%	95.08%	96.63%	96.55%

Experimental Results

Probabilistic Action Classification





Experimental Results

Detection Delay - Summary

Transition	Mean	Std	Median	Max	Min
Standing - Starting	57.98 ms	120.87 ms	50.00 ms	525.00 ms	-441.67 ms
Starting - Walking	-154.30 ms	183.66 ms	-208.33 ms	341.67 ms	-446.67 ms
Walking - Stopping	102.05 ms	157.86 ms	66.67 ms	416.67 ms	-450.00 ms
Stopping - Standing	89.84 ms	131.48 ms	58.33 ms	450.00 ms	-466.67 ms

Experimental Results

Video sequence showing prediction results

Experimental Results

UAH Prototype vehicle - DRIVERTIVE

Experimental Results

Experiments in UTAC proving ground (France) – BRAVE H2020 Project

Experimental Results

Experiments in UTAC proving ground (France) – BRAVE H2020 Project

Expected Contributions to Euro NCAP

- Recommendations to Euro NCAP WG on AD on how to measure performance of VRU prediction systems
- Enhanced performance in Euro NCAP tests by including predictions

Experimental Results

Experiments in UTAC proving ground (France) – BRAVE H2020 Project

Experimental Results

Experiments in UTAC proving ground (France) – BRAVE H2020 Project

Intelligent Interface with VRUs

GRAIL – GReen Assistant Interface Light

Intelligent Interface with VRUs

GRAIL – GReen Assistant Interface Light

Content

- Motivation
- Proposed approach
- Pedestrian pose measurement
- Dimensionality reduction and Prediction
- Experimental results
- **Conclusions and Future Work**

Conclusions and Future Work

Conclusions

- Accurate path pedestrian prediction is possible up to 1.0 s using body language traits and probabilistic dimensionality reduction techniques

-Pedestrian 3D pose and body joints are detected using Deep Learning and transformed later on to a latent space using B-GPDM

- Predictions are carried out in the latent space, yielding results with a potential accuracy of 15-20 cm at a time horizon of 1.0 s

Future Work

- Enhancement of prediction method using Deep Learning.

- Context-based action prediction using Probabilistic Graphical Models (Bayesian Networks) is under development, including:

    - Gaze direction, group behavior.

Enhanced Protection of Vulnerable Road Users

Thanks for your attention
Muito obrigado

Miguel Ángel Sotelo
President. IEEE ITS Society

miguel.sotelo@uah.es
Full Professor
University of Alcalá (UAH)
SPAIN