Tracking Known and Unknown Human Activites

John Darby, Baihua Li, Nicholas Costen,

Department of Computing and Mathematics, Manchester Metropolitan University

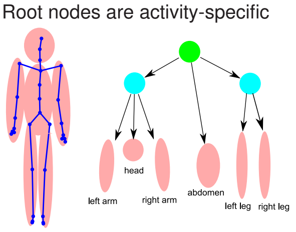
Introduction

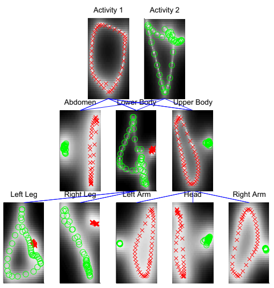
Constructing a pose space for analysis-by-synthesis

- **High-D Pose Space**
  - “Curse of dimensionality”
  - Need efficient search techniques
  - Partitioned Sampling [6], Annealed Particle Filter [2]
  - Potential to cope with any activity
- **Low-D Pose Space**
  - As few as 2-3 dimensions
  - Limited image evidence sufficient
  - Many available techniques
  - PCA [7], GP-LVM [8]
  - Activity specific

Hierarchical Models: H-GPLVM [5]

- **Learning**
  - Composed of GP-LVMs [4]
  - Represents high-D data through a low-D latent model and a non-linear GP mapping from latent space to data space
  - MoCap state space is partitioned between 5 nodes
  - Latent variables initialised through application of PCA to joint angles
  - Augmented by further latent models providing coordination
  - Non-leaf nodes model joint distribution over latent variables of children
  - Latent variables initialised through application of PCA to concatenated latent variables of children
  - Root nodes are activity-specific
- **Pose Generation**
  - Given a particular latent position in any node, the H-GPLVM defines Gaussian conditional distributions over
    1. the children (non-leaf nodes)
    2. the state space (leaf nodes)
  - These can be used to fully descend the hierarchy to the state space
  - Top level root nodes are akin to global activity models
  - Bottom level leaf nodes are akin to a flat part-based activity model
  - H-GPLVM can be used to produce novel poses depending on the extent to which coordination is respected





Pose Estimation

- Recover novel poses by ‘backing off’ down the hierarchy [5]
- Applying the models in the next level independently
- Particle-based approach
  1. initialised in the root nodes (globally coordinated training poses)
  2. terminating in leaf nodes (uncoordinated part-based poses)
- APF used to gradually introduce peaks in the cost function [2]
- Recombine particle coordinates for each latent space using crossover operator-type approach

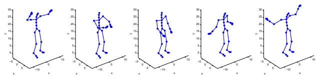
for $t = 1$ to T **do**
  Reinitialise from root data + noise: $\{(\mathbf{x}_{t,R}^{(n)})\}_{n=1}^{N}$
  for $r = R$ downto 1 **do**
    1. **Evaluate** weights $\pi_{t,r}^{(n)} = w_r(\mathbf{z}_t, \mathbf{x}_{t,r}^{(n)})$
    2. **Resample** B particles with likelihood $\propto \pi_{t,r}^{(n)}$ and with replacement
    3. **Back off** using mapping from latent coordinates to descend to next level
    4. **Recombine** particle coordinates for each node to form new particle set
    5. **Disperse** latent coordinates with noise term
  end for
  Calculate expected pose for visualisation $E(\mathbf{x}_t) = \sum_{n=1}^{N} \pi_{t,1}^{(n)} \mathbf{x}_{t,1}^{(n)}$.
end for

Figure 1: Pseudocode for pose estimation.

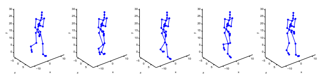
Weighting Function

- Compare joint locations in observation and hypotheses
- **MoCap**: squared 3D Euclidean distance
  - 15 joint locations on each body model
- **Monocular**: squared 2D Euclidean distance
  - 9 joint locations on hypothesised body model
  - 9 approximate joint locations in image
  - Found by 2D image-based tracker: WSL [3]
  - Manually initialised: few mouse clicks
  - Able to handle partial occlusions

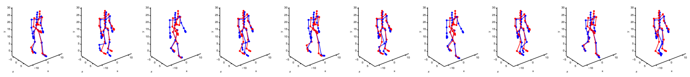
Results: MoCap Data



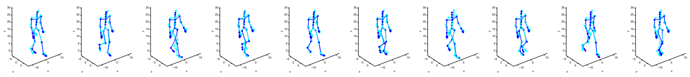
(a) Training data 1: *swinging arms*.



(b) Training data 2: *walking with hurt stomach*.



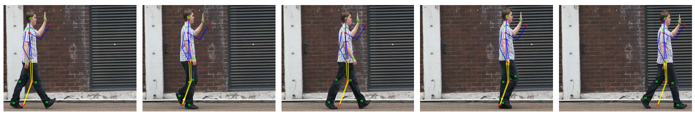
(c) GPDMs: the models are unable to generalise to novel poses. Particles oscillate between the best compromises in each latent space.



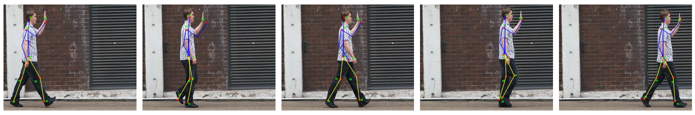
(d) H-GPLVM: back off and the addition of latent space noise allows the recovery of novel poses. Note the opposing swing of the arms.

Figure 2: MoCap training data [1] (a, b) and resulting pose estimation results for a *walking* sequence using GPDMs (c) and H-GPLVM (d).

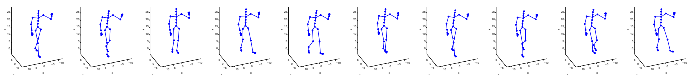
Results: Monocular Data



(a) GPDMs: neither latent space contains the pose. The waving hand is recovered at the expense of the legs.



(b) H-GPLVM: back off allows the combination of training data for accurate pose recovery.



(c) H-GPLVM: inferred 3D poses from a different view point.

Figure 3: Pose estimation results using 2D WSL joint tracks from a monocular *walking whilst waving* sequence: GPDMs (a), H-GPLVM (b). Training data is *slow walk/stride* and *stand and wave*.

Results: Occlusions

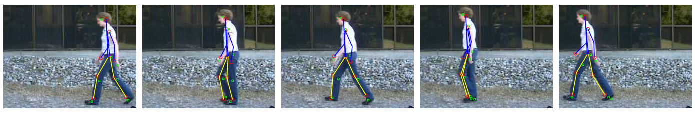


Figure 4: H-GPLVM: pose estimation for a walking sequence [7] using 2D WSL joint tracks. Position of occluded right arm is inferred from the visible upper body.

Conclusions

- **Discussion**
  - Outperforms global models for novel poses
  - By modelling correlations between nodes separately we can:
    1. Disregard them to recover novel poses (back off to leaf nodes)
    2. Respect them to handle occlusions (terminate descent early)
- **Future Work**
  - Find a complimentary set of “basis activities”
  - Final dispersion and resampling step in full state space
  - Make backing off a decision
  - Temporal model e.g. cluster and ascend

References

[1] CMU graphics lab motion capture database. http://mocap.cs.cmu.edu/.

[2] J. Deutscher and I. Reid. Articulated body motion capture by stochastic search. *IJCV*, 61(2):185–205, 2005.

[3] A. D. Jepson, D. J. Fleet, and T. F. El-Maraghi. Robust online appearance models for visual tracking. *PAMI*, 25(10):1296–1311, 2003.

[4] N. D. Lawrence. Probabilistic non-linear principal component analysis with Gaussian process latent variable models. *JMLR*, 6:1783–1816, 2005.

[5] N. D. Lawrence and A. J. Moore. Hierarchical Gaussian process latent variable models. In *ICML*, pages 481–488, 2007.

[6] J. MacCormick and M. Isard. Partitioned sampling, articulated objects, and interface-quality hand tracking. In *ECCV*, pages 3–19, 2000.

[7] H. Sidenbladh, M. J. Black, and L. Sigal. Implicit probabilistic models of human motion for synthesis and tracking. In *ECCV*, pages 784–800, 2002.

[8] R. Urtasun, D. J. Fleet, A. Hertzmann, and P. Fua. Priors for people tracking from small training sets. In *ICCV*, pages 403–410, 2005.

 Mail: n.costen@mmu.ac.uk WWW: http://www2.docm.mmu.ac.uk/STAFF/N.Costen