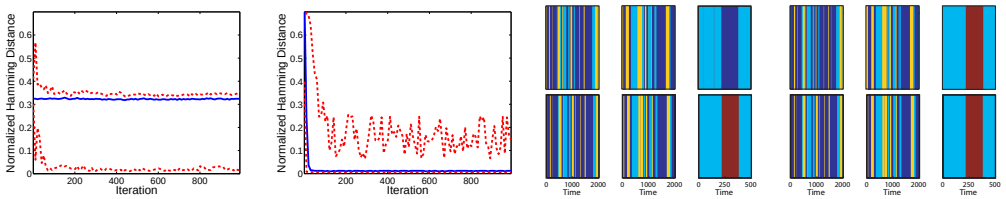

(a) (b) (c) (d)

Figure 3: (a)-(b) The 10th, 50th, and 90th Hamming distance quantiles for object 3 over 1000 trials for the HDP-AR-HMMs and BP-AR-HMM, respectively. (c)-(d) Examples of typical segmentations into behavior modes for the three objects at Gibbs iteration 1000 for the two models (top = estimate, bottom = truth).

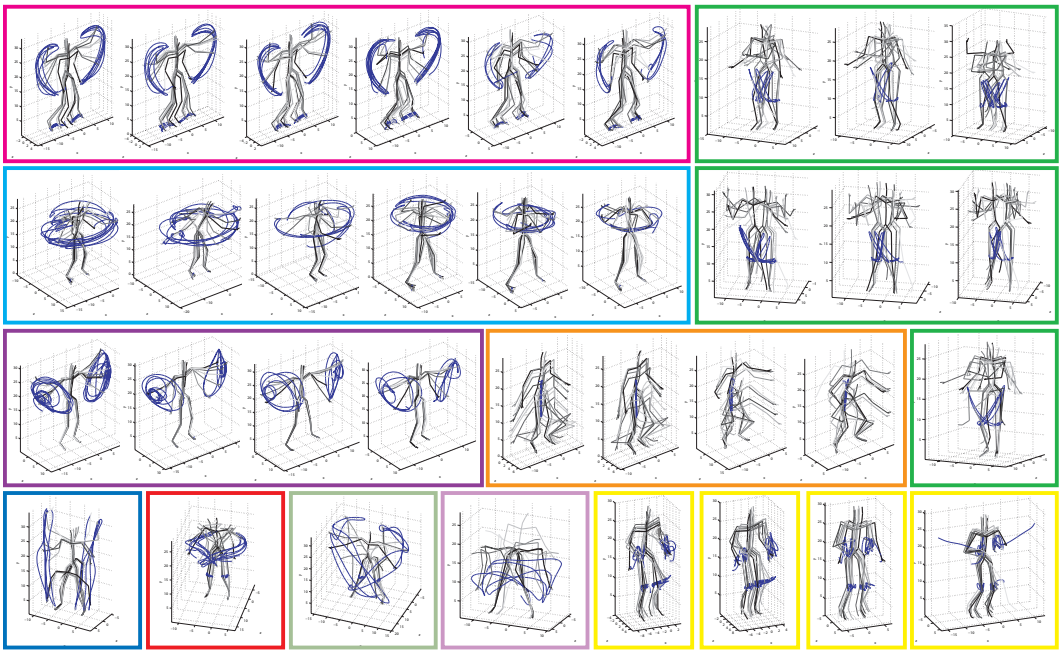

Figure 4: Each skeleton plot displays the trajectory of a learned contiguous segment of more than 2 seconds. To reduce the number of plots, we preprocessed the data to bridge segments separated by fewer than 300 msec. The boxes group segments categorized under the same feature label, with the color indicating the true feature label. Skeleton rendering done by modifications to Neil Lawrence's Matlab MoCap toolbox [13].

by $a_k \in \{-0.8, -0.4, 0.8\}$ and the third object used $a_k \in \{-0.3, 0.8\}$. The results shown in Fig. 3 indicate that the multiple HDP-AR-HMM model typically describes the third object using $a_k \in \{-0.4, 0.8\}$ since this assignment better matches the parameters defined by the other (lengthy) time series. These results reiterate that the feature model emphasizes choosing behaviors rather than assuming all objects are performing minor variations of the same dynamics.

For the experiments above, we placed a Gamma$(1, 1)$ prior on α and γ, and a Gamma$(100, 1)$ prior on κ. The gamma proposals used $\sigma_\gamma^2 = 1$ and $\sigma_\kappa^2 = 100$ while the MNIW prior was given $M = 0$, $K = 0.1 * I_d$, $n_0 = d + 2$, and S_0 set to 0.75 times the empirical variance of the joint set of first difference observations. At initialization, each time series was segmented into five contiguous blocks, with feature labels unique to that sequence.

6 Motion Capture Experiments

The linear dynamical system is a common model for describing simple human motion [11], and the more complicated SLDS has been successfully applied to the problem of human motion synthesis, classification, and visual tracking [17, 18]. Other approaches develop non-linear dynamical models using Gaussian processes [25] or based on a collection of binary latent features [20]. However, there has been little effort in jointly segmenting and identifying common dynamic behaviors amongst a set of *multiple* motion capture (MoCap) recordings of people performing various tasks. The BP-AR-HMM provides an ideal way of handling this problem. One benefit of the proposed model, versus the standard SLDS, is that it does not rely on manually specifying the set of possible behaviors.