A Distance for HMMs based on Aggregated Wasserstein Metric and State Registration

Yukun Chen, Jianbo Ye, and Jia Li
The Pennsylvania State University

PennState

ECCV'16

0. Motivation

KL based distance between HMMs: [Juang *et al.*]

$$d(\Lambda^{(1)},\Lambda^{(2)})\overset{\text{def}}{=}\int_{O_T}\frac{1}{T}\log\left(\frac{P_{\Lambda^{(1)}}(O_T)}{P_{\Lambda^{(2)}}(O_T)}\right)P_{\Lambda^{(1)}}(O_T)dO_T$$
$$=\lim_{T\to\infty}\frac{1}{T}\left(P_{\Lambda^{(1)}}(O_T)-P_{\Lambda^{(2)}}(O_T)\right)$$

	KL-based	MAW	IAW
rely on original seq. data	Yes, sometimes	No	No
need to generate long seq.	Yes, sometimes	No	Kind of
differentiation ability	Not good	Good	Best
Speed	Fast	Fast	Relative slow

Table: Summary

1. Introduction

Hidden Markov Model

- M states: $\{s_1, s_2, ..., s_M\}$
- Transition matrix $\mathbf{T}$: $\mathbf{T}_{i,j}\overset{\text{def}}{=}P(s_{t+1}=j|s_t=i)$
- Parameters of Gaussian probabilistic emission functions: $\mathcal{M}(\{\mu_i\}_{i=1}^M,\{\Sigma_i\}_{i=1}^M,\pi)$

We denote a HMM as:

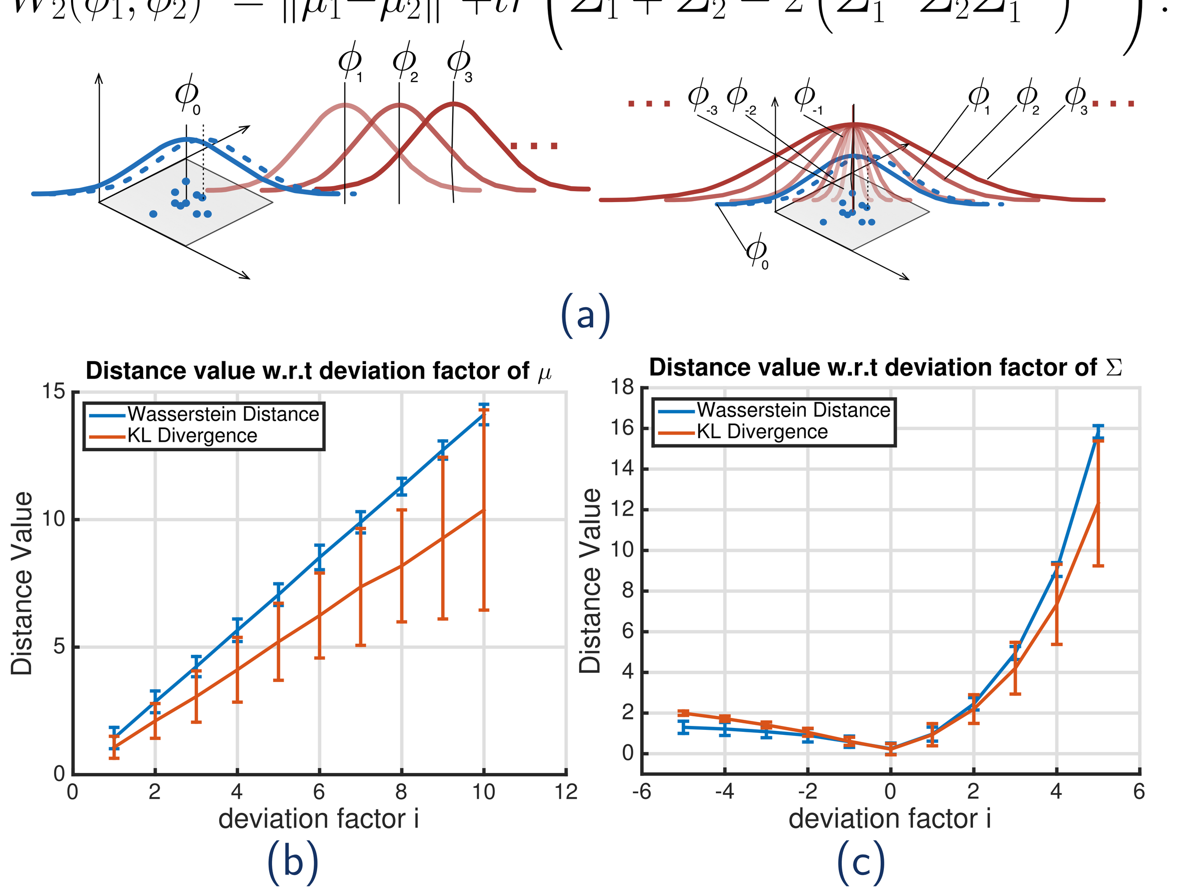
$$\Lambda(\mathbf{T},\mathcal{M})=\Lambda(\mathbf{T},\{\mu_i\}_{i=1}^M,\{\Sigma_i\}_{i=1}^M)$$

p-Wasserstein Distance

$$W_p(f,g)\overset{\text{def}}{=}\left[\inf_{\gamma\in\Pi(f,g)}\int_{\mathbb{R}^d\times\mathbb{R}^d}\|\mathbf{x}-\mathbf{y}\|^p d\gamma(\mathbf{x},\mathbf{y})\right]^{1/p}$$

If $p=2$, $\exists$ closed form solution for Gaussians:

$$W_2(\phi_1,\phi_2)^2=\|\mu_1-\mu_2\|^2+tr\left(\mathbf{\Sigma}_1+\mathbf{\Sigma}_2-2\left(\mathbf{\Sigma}_1^{1/2}\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{1/2}\right)^{1/2}\right).$$



(a) (b) (c)

Figure: (a) Experiment scheme for varying μ and varying Σ. A re-estimated $\widehat{\phi}_0$ is denoted as the dashed blue line. (b) (c) Mean estimates of $W_2(\widehat{\phi}_0,\phi_i)$ (blue) and $KL(\widehat{\phi}_0,\phi_i)$ (orange) and their 3σ confidence intervals w.r.t different Gaussian ϕ_i. (b) is for varying μ, and (c) is for varying Σ.

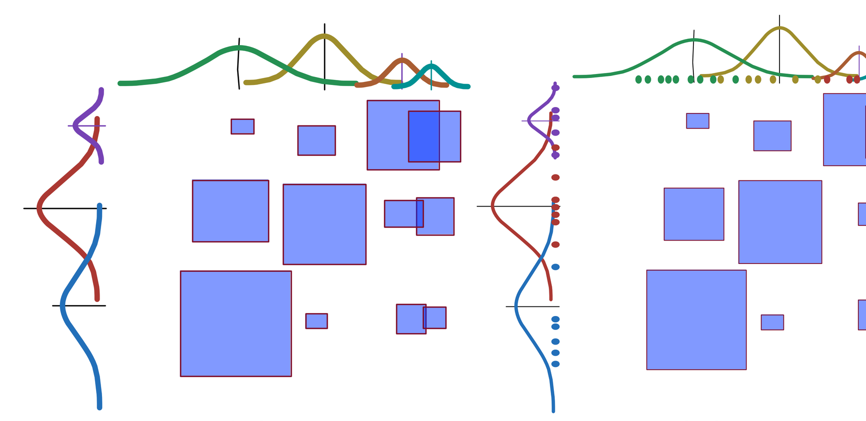
2. State Registration

Treat as an optimal transport problem:

$$\min_{\mathbf{W}\in\Pi(\pi_1,\pi_2)}\sum_{i=1}^{M_1}\sum_{j=1}^{M_2}w_{i,j}W_2(\phi_{1,i},\phi_{2,j})^p$$

where

$$\Pi(\pi_1,\pi_2)\overset{\text{def}}{=}\left\{\mathbf{W}\in\mathbb{R}^{M_1\times M_2}:\sum_{i=1}^{M_1}w_{i,j}=\pi_{2,j}, j=1,\ldots\right.$$
$$\left.\sum_{j=1}^{M_2}w_{i,j}=\pi_{1,i}, i=1,\ldots,M_1;\text{ and }\quad w_{i,j}\geq 0,\forall i,j\right\}$$



(a) (b)

Figure: Illustration of state registration. (a) for MAW , (b) for IAW

3. MAW and IAW

1) Difference between GMMs:

$$\widetilde{R}_p(\mathcal{M}_1,\mathcal{M}_2;\mathbf{W})^p\overset{\text{def}}{=}\sum_{i=1}^{M_1}\sum_{j=1}^{M_2}w_{i,j}W_2(\phi_{1,i},\phi_{2,j})^p$$

2) Difference between transition matrices:

$$\widetilde{\mathbf{T}}_2\overset{\text{def}}{=}\mathbf{W}_r\mathbf{T}_2\mathbf{W}_c^T\in\mathbb{R}^{M_1\times M_1},\ \widetilde{\mathbf{T}}_1\overset{\text{def}}{=}\mathbf{W}_c^T\mathbf{T}_1\mathbf{W}_r\in\mathbb{R}^{M_2\times M_2}.$$

$$d_T(\mathbf{T}_1,\widetilde{\mathbf{T}}_2)^p\overset{\text{def}}{=}\sum_{i=1}^{M_1}\pi_{1,i}\widetilde{W}_p\left(\mathcal{M}_1^{(i)}|_{\mathbf{T}_1(i,:)},\mathcal{M}_1^{(i)}|_{\widetilde{\mathbf{T}}_2(i,:)}\right)^p$$
$$d_T(\mathbf{T}_2,\widetilde{\mathbf{T}}_1)^p\overset{\text{def}}{=}\sum_{i=1}^{M_2}\pi_{2,i}\widetilde{W}_p\left(\mathcal{M}_2^{(i)}|_{\mathbf{T}_2(i,:)},\mathcal{M}_2^{(i)}|_{\widetilde{\mathbf{T}}_1(i,:)}\right)^p$$

$$D_p(\mathbf{T}_1,\mathbf{T}_2:\mathbf{W})^p\overset{\text{def}}{=}d_T(\mathbf{T}_1,\widetilde{\mathbf{T}}_2)^p+d_T(\mathbf{T}_2,\widetilde{\mathbf{T}}_1)^p$$

Aggregated Wasserstein:(MAW/IAW)

$$AW(\Lambda_1,\Lambda_2)\overset{\text{def}}{=}$$
$$(1-\alpha)\widetilde{R}_p(\mathcal{M}_1,\mathcal{M}_2;\mathbf{W})+\alpha D_p(\mathbf{T}_1,\mathbf{T}_2:\mathbf{W})$$

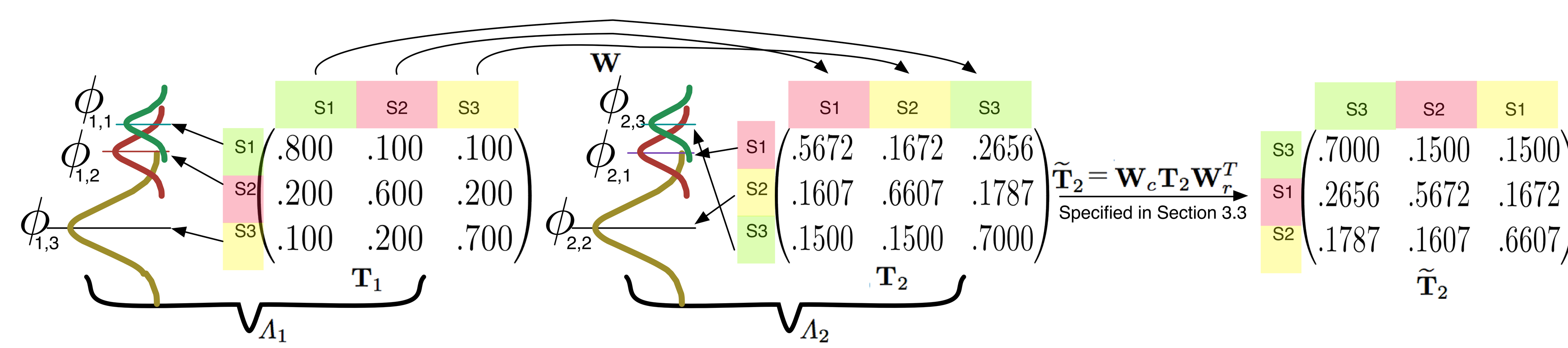


Figure: A simple registration example about how $\mathbf{T}_2$ in Λ_2 is registered towards Λ_1 such that it can be compared with $\mathbf{T}_1$ in Λ_1. For this example, $\mathbf{W}$ encodes a "hard matching" between states in Λ_1 and Λ_2

4. Experiments

Synthetic Data: (1-NN Retrieval)

Exp. index	deviation step	$\tilde{\mu}$	$\tilde{\Sigma}$	$\mathbf{T}$
1	$\Delta\mu=0.2$, 0.4, 0.6	$\left\{\begin{pmatrix}2+i\Delta\mu\\2+i\Delta\mu\end{pmatrix},\begin{pmatrix}5+i\Delta\mu\\5+i\Delta\mu\end{pmatrix}\Big\vert i=1,2,3,4,5\right\}$	$\left\{\begin{pmatrix}1&0\\0&1\end{pmatrix},\begin{pmatrix}1&0\\0&1\end{pmatrix}\right\}$	$\begin{pmatrix}0.8&0.2\\0.2&0.8\end{pmatrix}$
2	$\Delta\sigma=0.2$, 0.4, 0.6	$\left\{\begin{pmatrix}2\\2\end{pmatrix},\begin{pmatrix}5\\5\end{pmatrix}\right\}$	$\{[0.2\cdot exp(i\Delta\sigma\cdot\mathbf{S}), 0.2\cdot exp(i\Delta\sigma\cdot\mathbf{S})]\vert i=1,2,3,4,5\}$, $\mathbf{S}=rand(2)$	$\begin{pmatrix}0.8&0.2\\0.2&0.8\end{pmatrix}$
3	$\Delta t=0.2$, 0.4, 0.6	$\left\{\begin{pmatrix}2\\2\end{pmatrix},\begin{pmatrix}5\\5\end{pmatrix}\right\}$	$\left\{\begin{pmatrix}1&0\\0&1\end{pmatrix},\begin{pmatrix}1&0\\0&1\end{pmatrix}\right\}$	$\{\Delta t\cdot\mathbf{S}+(1-\Delta t)\cdot\mathbf{T}_i\vert \mathbf{T}_i[j,:]\sim Dirichlet(10\cdot\mathbf{S}[j,:]), i=1,2,3,4,5\}$, $\mathbf{S}=\begin{pmatrix}0.8&0.2\\0.2&0.8\end{pmatrix}$

Table: Summary of the parameters setup for parameter perturbation experiments. $rand(2)$ here means random matrix of dimension 2 by 2. $Dirichlet(\vec{x})$ here means generating samples from Direchlet distribution with parameter $\vec{x}$.

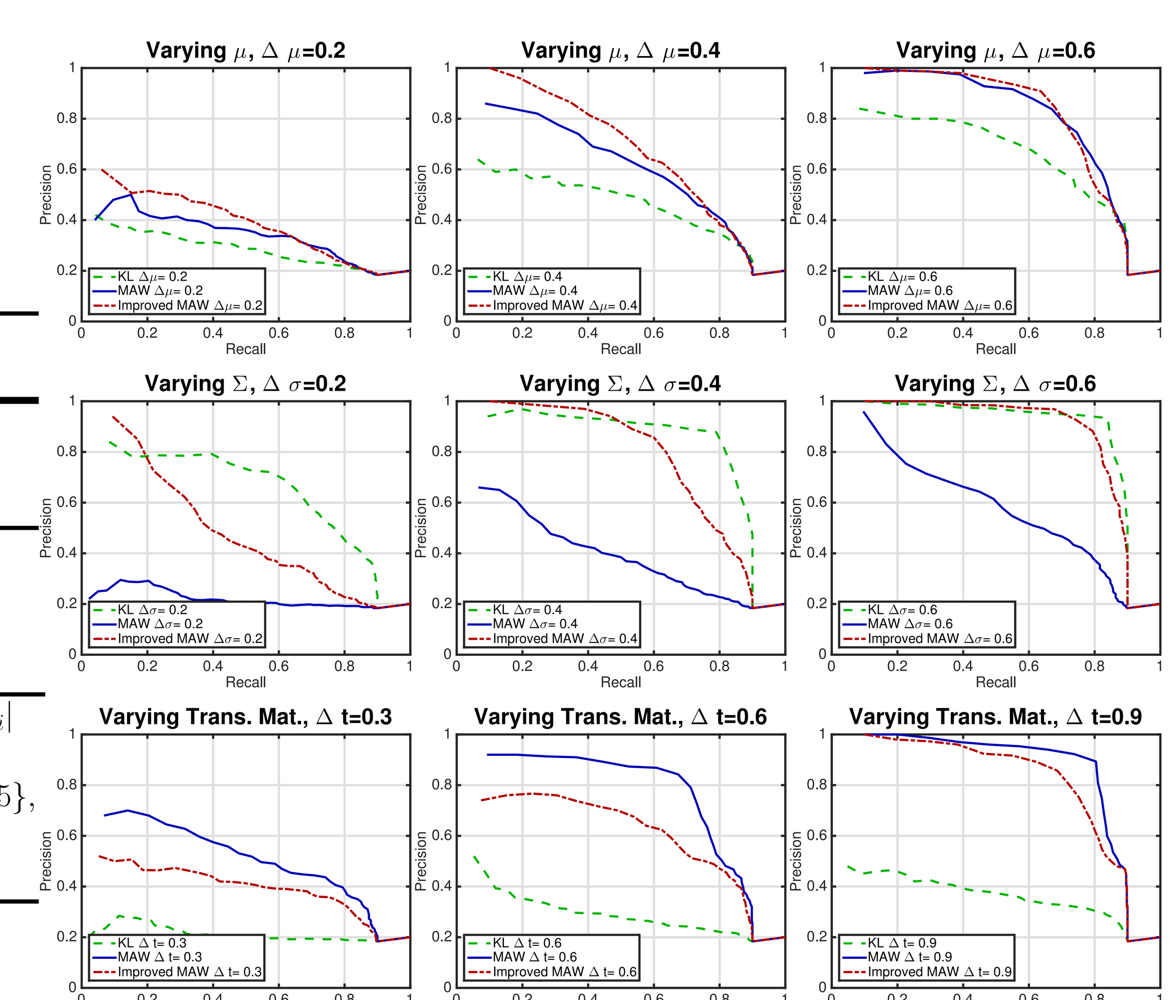


Figure: Precision-recall plot for the study to compare KL, MAW and IAW's sensitivity to the perturbation of

4. Experiments (Cont.)

CMU Mocap Data [a]:

Figure: Visualization of CMU motion capture data. Left: Jump. Right: Walk

1). 1-NN Retrieval:

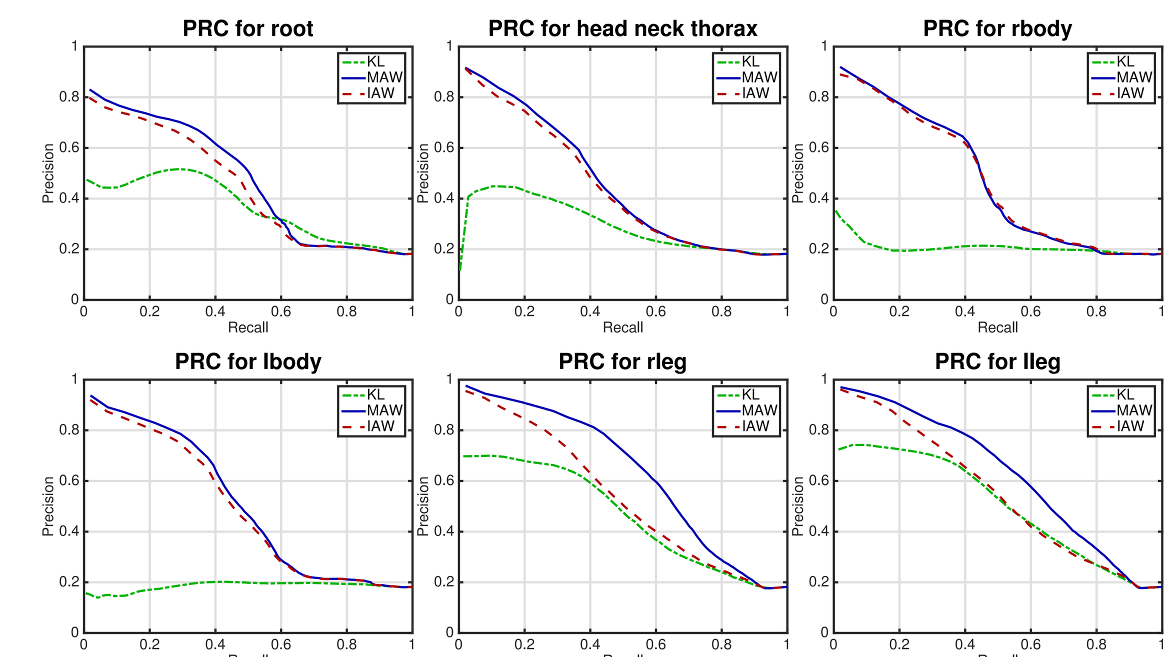
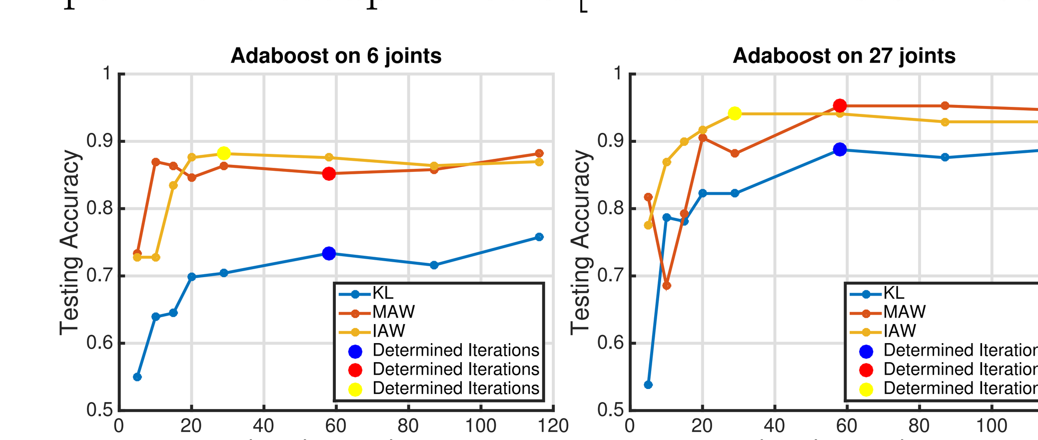


Figure: Precision Recall Plot for Motion Retrieval. The plot for 6 joint-groups, i.e. $root_{12}$, $head\_neck\_thorax_{12}$, $rbody_{12}$, $lbody_{12}$, $rleg_6$, $lleg_6$, are displayed separately.

2). Classification by Adaboost:

Experiment setup follows [Lv *et al.* ECCV'2006]



(a) (b)

Figure: Testing accuracies w.r.t iteration number of Adaboost (number of weak classifiers selected). (a) Motion Classification by Adaboost on 6 joints. (b) Motion Classification by Adaboost on 27 joints. The iteration number means the number of features incrementally acquired in Adaboost.

5. References

- Juang, B.H.F., Rabiner, L.R.: A probabilistic distance measure for hidden Markov models. AT&T Technical Journal (1985)
- Lv, F., Nevatia, R.: Recognition and segmentation of 3-d human action using hmm and multi-class adaboost. (ECCV 2006).

[a] http://mocap.cs.cmu.edu