Supplementary Material
No Bias Left Behind: Covariate Shift Adaptation for Discriminative 3D Pose Estimation

Makoto Yamada[1], Leonid Sigal[2], Michalis Raptis[2]

[1]NTT Communication Science Laboratories
[2]Disney Research Pittsburgh
yamada@cs.titech.ac.jp
{lsigal mraptis}@disneyresearch.com

1 Derivation in Importance Weight Estimation

The parameters $\boldsymbol{\theta}$ in the model $w_\alpha(\boldsymbol{x};\boldsymbol{\theta})$ are determined so that the following expected squared-error J is minimized:

$$
\begin{aligned}
J(\boldsymbol{\theta}) &= \frac{1}{2}\mathbb{E}_{q_\alpha(\boldsymbol{x})}\left[(w_\alpha(\boldsymbol{x};\boldsymbol{\theta}) - w_\alpha(\boldsymbol{x}))^2\right] \\
&= \frac{1}{2}\mathbb{E}_{q_\alpha(\boldsymbol{x})}\left[w_\alpha(\boldsymbol{x};\boldsymbol{\theta})^2 - 2w_\alpha(\boldsymbol{x};\boldsymbol{\theta})w_\alpha(\boldsymbol{x}) + w_\alpha(\boldsymbol{x})^2\right] \\
&= \frac{1}{2}\mathbb{E}_{q_\alpha(\boldsymbol{x})}[w_\alpha(\boldsymbol{x};\boldsymbol{\theta})^2] - \mathbb{E}_{q_\alpha(\boldsymbol{x})}[w_\alpha(\boldsymbol{x};\boldsymbol{\theta})w_\alpha(\boldsymbol{x})] + \text{Const.} \\
&= \frac{(1-\alpha)}{2}\mathbb{E}_{p_{\text{te}}(\boldsymbol{x})}\left[w_\alpha(\boldsymbol{x};\boldsymbol{\theta})^2\right] + \frac{\alpha}{2}\mathbb{E}_{p_{\text{tr}}(\boldsymbol{x})}\left[w_\alpha(\boldsymbol{x};\boldsymbol{\theta})^2\right] - \mathbb{E}_{p_{\text{te}}(\boldsymbol{x})}[w_\alpha(\boldsymbol{x};\boldsymbol{\theta})] + \text{Const.},
\end{aligned}
$$

where $q_\alpha(\boldsymbol{x}) = (1-\alpha)p_{\text{te}}(\boldsymbol{x}) + \alpha p_{\text{tr}}(\boldsymbol{x})$, and we used $w_\alpha(\boldsymbol{x})q_\alpha(\boldsymbol{x}) = p_{\text{te}}(\boldsymbol{x})$.

(a) Data samples   (b) KR (MSE=0.021)   (c) IWTGP (MSE=0.002)
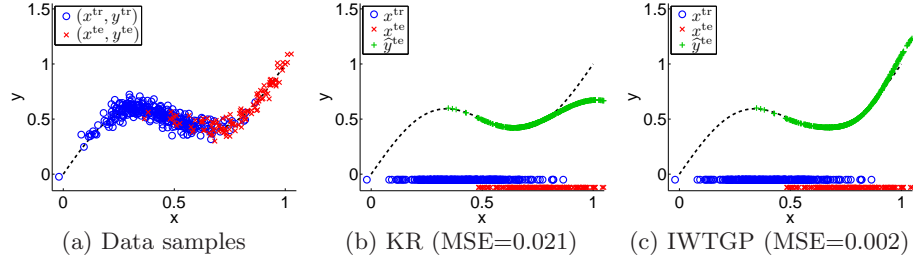
Fig. 1. Predicted outputs y by KR (b) and IWKR (c) under covariate shift in green. (a) Samples from the model $y = x + 0.3\sin(2\pi x) + e$ where $e \sim \mathcal{N}(0, 0.05^2)$; $\circ$ and $\times$ are training and test samples respectively (for clarity we also illustrate marginals $p_{\mathrm{tr}}(\boldsymbol{x})$ and $p_{\mathrm{te}}(\boldsymbol{x})$ in (b) and (c) bottom). Note that the input-output test samples are not used in the training of KR and the output test samples are not used in the training of IWKR, they are plotted in the figure for illustration purposes.
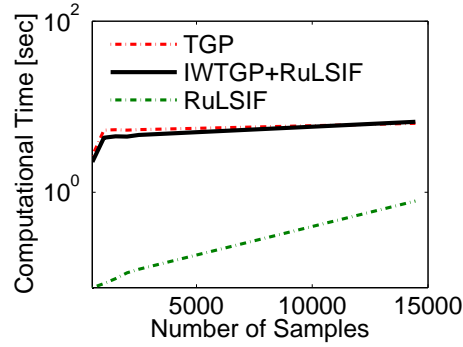

Fig. 2. Computational time comparison. Mean computational time of TGP, IWTGP + RuLSIF, and RuLSIF with respect to the number of training samples. The horizontal axis denotes the number of training samples, and the vertical axis denotes the computational time (sec). In this experiment, we use "Boxing" sequence of subject1 for testing, where the number of frames is 251.
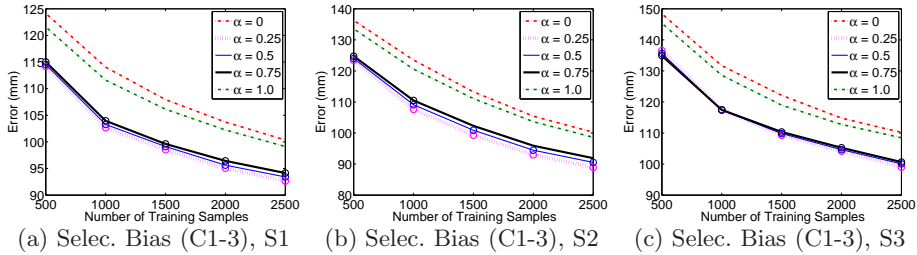

(a) Selec. Bias (C1-3), S1   (b) Selec. Bias (C1-3), S2   (c) Selec. Bias (C1-3), S3

Fig. 3. Performance comparison of IWTGP with respect to the α parameter on HUMANEVA-I dataset illustrated as a function of the number of training samples; we averaged the error over all motions for each subject. The best method achieving the smallest mean error and comparable methods according to the paired *t-test* at the significance level 5% are specified by '$\circ$'.

Table 1. Performance on the entire HUMANEVA-I dataset; we average errors over all motions and frames.

| Transfer Type	Subject Train	Subject Test	Motion	IWTGP	TGP	IWKR	KR	WkNN
Selection Bias (C1)	S1,S2,S3	S1	Walking	**40.4**	41.1	63.6	69.6	57.2
			Jogging	**51.7**	52.0	74.6	85.3	75.6
			Gestures	**10.5**	11.1	15.9	18.3	11.8
			Box	**39.1**	40.0	55.5	62.4	50.1
			ThrowCatch	**129.1**	131.5	145.7	165.2	156.6
			Average	**54.2**	55.1	71.1	80.1	70.2
	S1,S2,S3	S2	Walking	**24.8**	26.5	44.5	55.1	41.7
			Jogging	**39.3**	40.2	55.6	63.1	60.0
			Gestures	65.2	**65.0**	77.8	83.0	72.4
			Box	**59.9**	60.5	76.2	85.6	91.7
			ThrowCatch	74.2	**74.0**	83.9	90.6	91.9
			Average	**52.5**	53.2	67.6	75.5	71.5
	S1,S2,S3	S3	Walking	**57.7**	59.2	80.6	91.0	80.9
			Jogging	**34.2**	35.2	53.9	64.7	58.7
			Gestures	72.7	**70.6**	91.9	109.5	82.7
			Box	**65.3**	66.8	73.8	78.6	67.6
			ThrowCatch	—	—	—	—	—
			Average	**57.5**	57.9	75.1	86.0	72.5
Selection Bias (C1-3)	S1,S2,S3	S1	Walking	**92.4**	96.3	120.4	132.9	114.5
			Jogging	**122.9**	123.5	130.8	142.6	130.4
			Gestures	**11.2**	12.0	26.3	43.7	12.0
			Box	**42.0**	43.4	68.5	91.0	54.2
			ThrowCatch	**141.2**	144.1	163.5	189.1	160.3
			Average	**81.9**	83.9	101.9	119.9	94.3
	S1,S2,S3	S2	Walking	**36.3**	39.9	85.9	111.4	68.6
			Jogging	**94.3**	97.1	127.0	141.4	146.3
			Gestures	**89.6**	91.0	103.7	115.4	97.8
			Box	**60.8**	62.7	89.8	108.4	83.0
			ThrowCatch	**82.4**	84.9	107.3	123.2	107.6
			Average	**72.7**	75.1	102.7	120.0	100.7
	S1,S2,S3	S3	Walking	**98.6**	104.5	137.2	154.2	132.8
			Jogging	**64.4**	70.5	121.4	144.9	127.7
			Gestures	**87.9**	110.4	102.6	139.8	106.0
			Box	**58.0**	59.1	85.5	100.3	75.2
			ThrowCatch	—	—	—	—	—
			Average	**77.2**	86.1	111.7	134.8	110.4
Subject Transfer (C1)	S2,S3	S1	Walking	**94.9**	98.3	115.3	143.4	95.6
			Jogging	106.0	110.6	115.6	135.6	**104.6**
			Gestures	**86.3**	87.9	121.4	187.0	100.9
			Box	151.6	148.5	**142.5**	156.8	144.0
			ThrowCatch	191.9	**189.4**	191.8	219.0	196.0
			Average	**126.2**	126.9	137.3	168.4	128.2
	S1,S3	S2	Walking	**108.9**	**108.9**	124.5	127.2	126.9
			Jogging	92.1	**91.1**	117.4	121.3	141.8
			Gestures	96.3	**94.8**	110.1	128.5	100.8
			Box	177.6	178.1	168.5	184.1	**168.8**
			ThrowCatch	**108.4**	109.9	132.2	146.5	114.6
			Average	116.7	**116.6**	130.5	141.5	130.6
	S1,S2	S3	Walking	**130.7**	132.7	150.1	192.7	133.6
			Jogging	**116.9**	120.2	147.3	193.4	139.6
			Gestures	**133.4**	197.9	186.5	191.2	140.1
			Box	179.2	188.0	189.7	259.2	**168.5**
			ThrowCatch	—	—	—	—	—
			Average	**140.0**	159.7	168.4	209.1	145.5
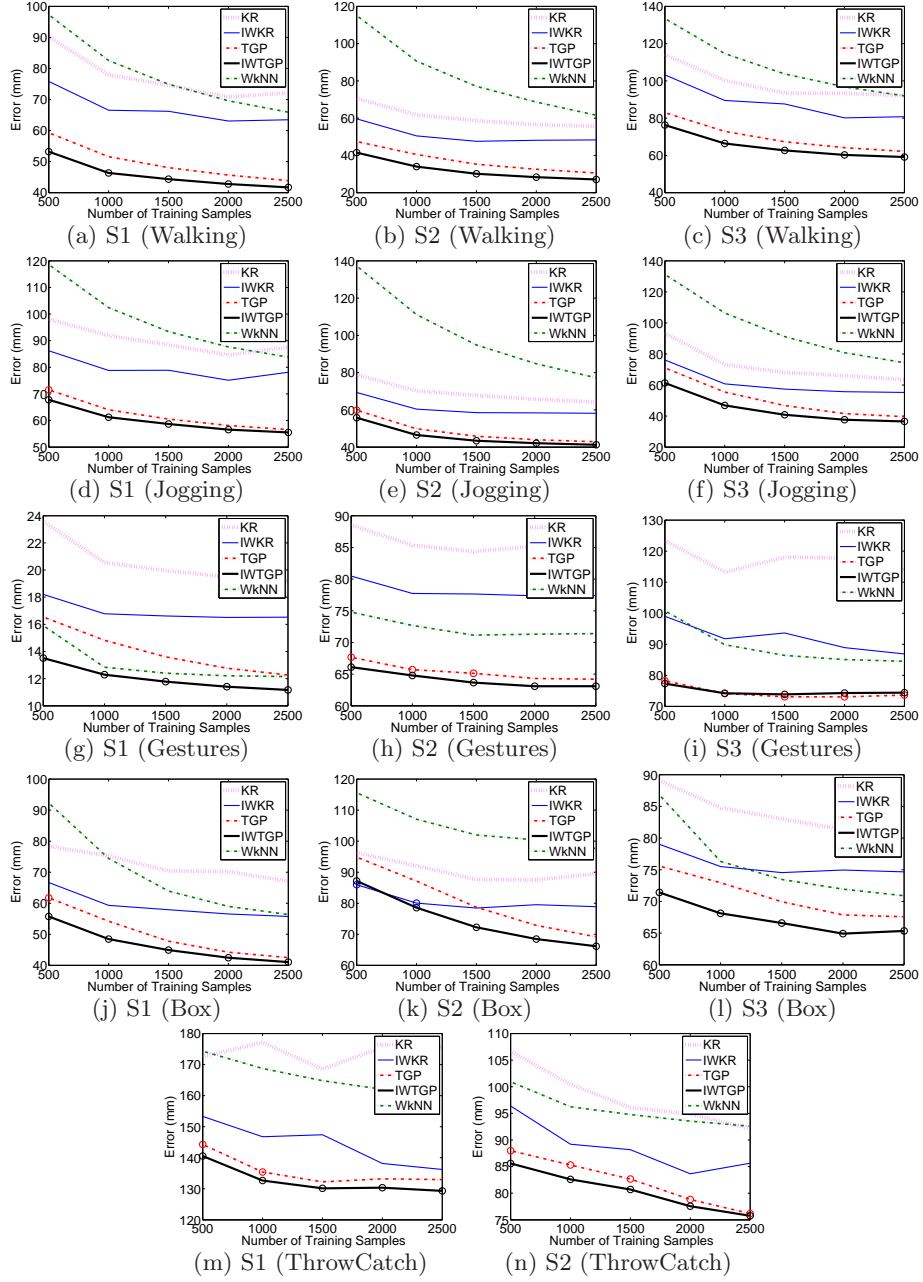
Fig. 4. Results of selection bias (C1). Mean error for HUMANEVA-I data with respect to the number of training samples for each subject and each motion. The horizontal axis denotes the number of training samples, and the vertical axis denotes the mean error. The best method achieving the smallest mean error and comparable methods according to the paired *t-test* at the significance level 5% are specified by '∘'.
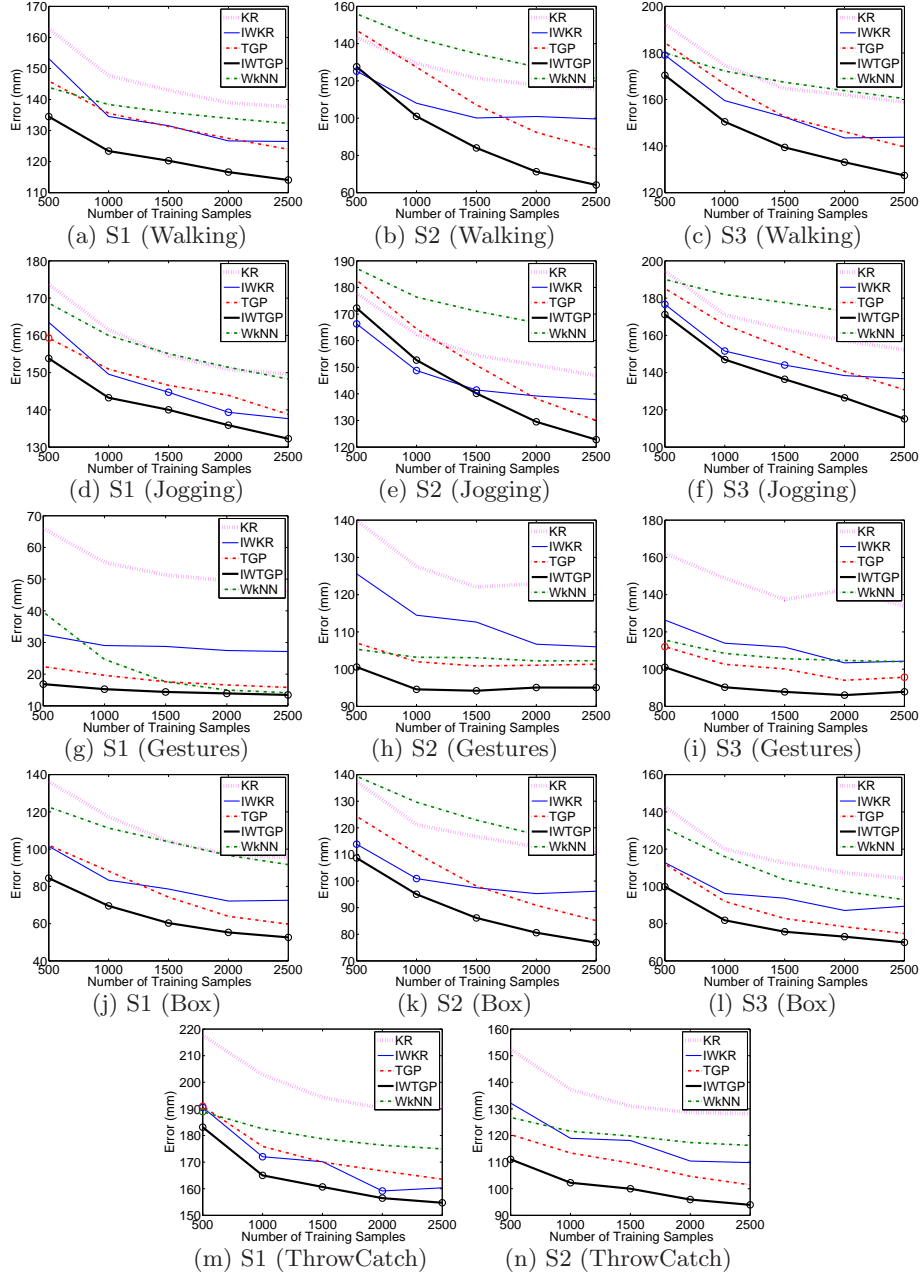
Fig. 5. Results of selection bias (C1-3). Mean error for HUMANEVA-I data with respect to the number of training samples for each subject and each motion. The horizontal axis denotes the number of training samples, and the vertical axis denotes the mean error. The best method achieving the smallest mean error and comparable methods according to the paired *t-test* at the significance level 5% are specified by '∘'.
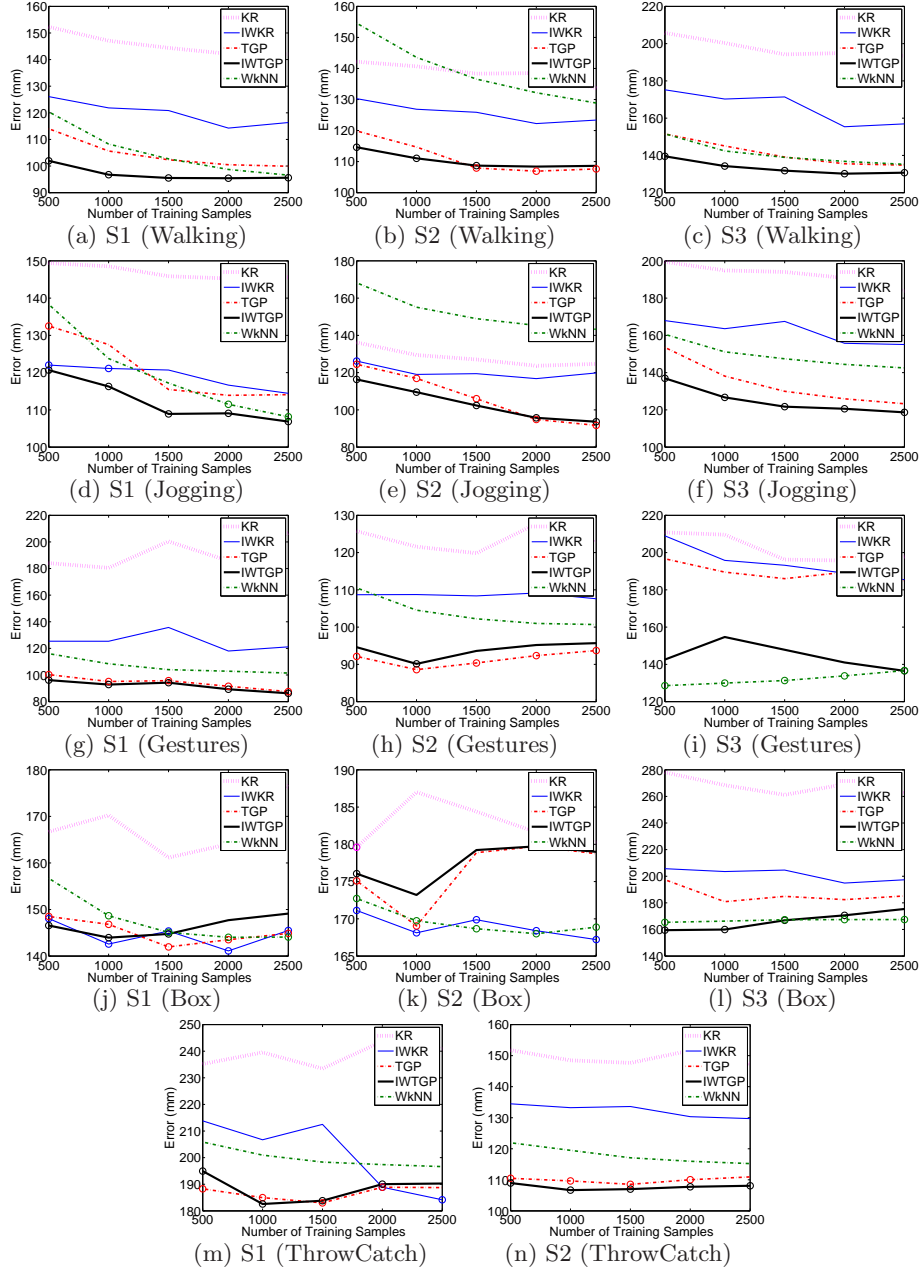
Fig. 6. Results of subject transfer. Mean error for HumanEva-I data with respect to the number of training samples for each subject and each motion. The horizontal axis denotes the number of training samples, and the vertical axis denotes the mean error. The best method achieving the smallest mean error and comparable methods according to the paired *t-test* at the significance level 5% are specified by 'o'.