Supplementary Material for BPCCA

Detailed Row-wise Model Derivations for Estimating $\boldsymbol{\theta}_r$

Based on the row-wise two-stage model (7), the conditional distribution of $\mathbf{X}^{(v)}$ given $\mathbf{Y}^{(v,r)}$ is $\mathbf{X}^{(v)}|\mathbf{Y}^{(v,r)} \sim \mathcal{N}_{d_v^c,d_v^r}(\mathbf{C}^{(v)}\mathbf{Y}^{(v,r)}, \boldsymbol{\Sigma}_c^{(v)}, \boldsymbol{\Psi}_r^{(v)})$. It can be rewritten in a vector form: $\boldsymbol{x}^{(v,r)}|\boldsymbol{y}^{(v,r)} \sim \mathcal{N}(\hat{\mathbf{C}}^{(v)}\boldsymbol{y}^{(v,r)}, \boldsymbol{\Psi}_r^{(v)} \otimes \boldsymbol{\Sigma}_c^{(v)})$, where $\boldsymbol{x}^{(v,r)} = \text{vec}(\mathbf{X}^{(v)})$, $\boldsymbol{y}^{(v,r)} = \text{vec}(\mathbf{Y}^{(v,r)})$, and $\hat{\mathbf{C}}^{(v)} = \mathbf{I} \otimes \mathbf{C}^{(v)}$.

Vector-based combination for $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$: Since $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are independent given $\mathbf{Y}^{(1,r)}$ and $\mathbf{Y}^{(2,r)}$, the two-view observations $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ can be combined via vector-based concatenation as follows:

$$\boldsymbol{x}^r|\boldsymbol{y}^r \sim \mathcal{N}(\hat{\mathbf{C}}\boldsymbol{y}^r, \mathbf{L}_r), \tag{21}$$

where $\boldsymbol{x}^r = [{\boldsymbol{x}^{(1,r)}}^\top, {\boldsymbol{x}^{(2,r)}}^\top]^\top$, $\boldsymbol{y}^r = [{\boldsymbol{y}^{(1,r)}}^\top, {\boldsymbol{y}^{(2,r)}}^\top]^\top$, $\hat{\mathbf{C}} = \text{blkdiag}(\hat{\mathbf{C}}^{(1)}, \hat{\mathbf{C}}^{(2)})$, and $\mathbf{L}_r = \text{blkdiag}(\boldsymbol{\Psi}_r^{(1)} \otimes \boldsymbol{\Sigma}_c^{(1)}, \boldsymbol{\Psi}_r^{(2)} \otimes \boldsymbol{\Sigma}_c^{(2)})$.

Matrix-based combination for $\mathbf{Y}^{(1,r)}$ and $\mathbf{Y}^{(2,r)}$: Based on (7), the two-view intermediate matrices $\mathbf{Y}^{(1,r)}$ and $\mathbf{Y}^{(2,r)}$ can be naturally combined via matrix-based concatenation as follows:

$$\mathbf{Y}^r|\mathbf{Z} \sim \mathcal{N}_{d_1^r+d_2^r,q^c}(\mathbf{R}\mathbf{Z}^\top, \boldsymbol{\Sigma}_r, \mathbf{I}), \tag{22}$$

where $\mathbf{Y}^r = [\mathbf{Y}^{(1,r)}, \mathbf{Y}^{(2,r)}]^\top$, $\mathbf{R} = [{\mathbf{R}^{(1)}}^\top, {\mathbf{R}^{(2)}}^\top]^\top$, and $\boldsymbol{\Sigma}_r = \text{blkdiag}(\boldsymbol{\Sigma}_r^{(1)}, \boldsymbol{\Sigma}_r^{(2)})$. Then, we can derive other distributions involved in $\mathbf{Y}^r$ and $\boldsymbol{y}^r$ as follows:

$$\mathbf{Y}^r \sim \mathcal{N}_{d_1^r+d_2^r,q^c}(\mathbf{0}, \boldsymbol{\Psi}_r, \mathbf{I}), \tag{23}$$

$$\boldsymbol{y}^r \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_r \otimes \mathbf{I}), \tag{24}$$

$$\mathbf{Z}|\mathbf{Y}^r \sim \mathcal{N}_{q^c,q^r}({\mathbf{Y}^r}^\top \boldsymbol{\Sigma}_r^{-1}\mathbf{R}\mathbf{M}_r, \mathbf{I}, \mathbf{M}_r), \tag{25}$$

$$\boldsymbol{y}^r|\boldsymbol{x}^r \sim \mathcal{N}(\boldsymbol{\Pi}_r\hat{\mathbf{C}}^\top\mathbf{L}_r^{-1}\boldsymbol{x}^r, \boldsymbol{\Pi}_r), \tag{26}$$

where $\boldsymbol{\Psi}_r = \mathbf{R}\mathbf{R}^\top + \boldsymbol{\Sigma}_r$, $\mathbf{M}_r = (\mathbf{R}^\top\boldsymbol{\Sigma}_r^{-1}\mathbf{R} + \mathbf{I})^{-1}$, and $\boldsymbol{\Pi}_r = (\boldsymbol{\Psi}_r^{-1} \otimes \mathbf{I} + \hat{\mathbf{C}}^\top\mathbf{L}_r^{-1}\hat{\mathbf{C}})^{-1}$.

Row-wise parameter estimation: If the intermediate matrices $\{\mathbf{Y}_n^r\}_{n=1}^N$ are observed, $\boldsymbol{\theta}_r$ can be solved via the EM algorithm by maximizing $\mathcal{L}_r(\boldsymbol{\theta}_r) = \sum_{n=1}^N \ln p(\mathbf{Y}_n^r, \mathbf{Z}_n|\boldsymbol{\theta}_r) = \sum_{n=1}^N \ln p(\mathbf{Y}_n^r|\mathbf{Z}_n, \boldsymbol{\theta}_r)p(\mathbf{Z}_n)$. Therefore, we can obtain the statistics of $\{\mathbf{Y}_n^c\}_{n=1}^N$ via their maximum posteriori estimations according to $p(\boldsymbol{y}^c|\boldsymbol{x}^c)$ (13), and turn to maximize the expectation of $\mathcal{L}_c(\boldsymbol{\theta}_c)$ w.r.t. $p(\boldsymbol{y}^c|\boldsymbol{x}^c)$ instead of the complicated log-likelihood $\mathcal{L}(\boldsymbol{\theta}_c, \boldsymbol{\theta}_r)$ for close-form solutions.

In the *E step*, we take the expectation of $\mathcal{L}_r(\boldsymbol{\theta}_r)$ w.r.t. $p(\mathbf{Z}, \mathbf{Y}^r|\boldsymbol{x}^r) = p(\mathbf{Z}|\mathbf{Y}^r)p(\boldsymbol{y}^r|\boldsymbol{x}^r)$ and obtain

$$\begin{aligned}\mathcal{Q}_r(\boldsymbol{\theta}_r) = -\frac{1}{2}\sum_{n=1}^N \Big\{ & q^c \ln|\boldsymbol{\Sigma}_r| + \text{tr}(\langle \boldsymbol{\Sigma}_r^{-1}(\mathbf{Y}_n^r{\mathbf{Y}_n^r}^\top \\ & + \mathbf{R}\mathbf{Z}_n^\top\mathbf{Z}_n\mathbf{R}^\top - \mathbf{Y}_n^r\mathbf{Z}_n\mathbf{R}^\top - \mathbf{R}\mathbf{Z}_n^\top{\mathbf{Y}_n^r}^\top)\rangle_r) \Big\},\end{aligned} \tag{27}$$

where terms of $p(\mathbf{Z}_n)$ have been omitted as a constant, and $\langle\cdot\rangle_r$ denotes the expectation $\mathbb{E}[\mathbb{E}[\cdot|\mathbf{Y}^r]|\boldsymbol{x}^r]$ w.r.t. $p(\mathbf{Z}|\mathbf{Y}^r)$ and $p(\boldsymbol{y}^r|\boldsymbol{x}^r)$ correspondingly. According to (26), the required expectation is given by

$$\mathbb{E}[\mathbf{Y}_n^r{\mathbf{Y}_n^r}^\top|\boldsymbol{x}_n^r] = \text{tr}_{q^c}(\mathbb{E}[\boldsymbol{y}_n^r{\boldsymbol{y}_n^r}^\top|\boldsymbol{x}_n^r]), \tag{28}$$

where $\mathbb{E}[\boldsymbol{y}_n^r{\boldsymbol{y}_n^r}^\top|\boldsymbol{x}_n^r] = \boldsymbol{\Pi}_r + \boldsymbol{\Pi}_r\hat{\mathbf{C}}^\top\mathbf{L}_r^{-1}\boldsymbol{x}_n^r{\boldsymbol{x}_n^r}^\top\mathbf{L}_r^{-\top}\hat{\mathbf{C}}\boldsymbol{\Pi}_r$ is a $(d_1^r + d_2^r)q^c \times (d_1^r + d_2^r)q^c$ block matrix with $q^c \times q^c$ submatrices.

In the *M step*, we maximize $\mathcal{Q}_r(\boldsymbol{\theta}_r)$ (27) w.r.t. $\boldsymbol{\theta}_r$, which leads to the following solutions:

$$\tilde{\mathbf{R}} = \left[\sum_{n=1}^N \langle \mathbf{Y}_n^r\mathbf{Z}_n\rangle_r\right]\left[\sum_{n=1}^N \langle {\mathbf{Z}_n}^\top\mathbf{Z}_n\rangle_r\right]^{-1}, \tag{29}$$

$$\tilde{\boldsymbol{\Sigma}}_r^{(v)} = \frac{1}{Nq^c}\sum_{n=1}^N \langle \mathbf{H}_n^{(v,r)}{\mathbf{H}_n^{(v,r)}}^\top\rangle_r. \tag{30}$$

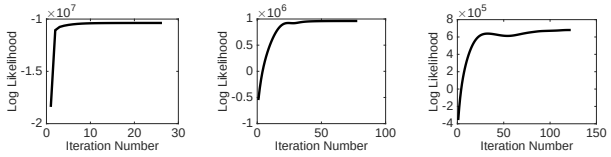
where $\langle \mathbf{Y}_n^r\mathbf{Z}_n\rangle_r = \mathbb{E}[\mathbf{Y}_n^r{\mathbf{Y}_n^r}^\top|\boldsymbol{x}_n^r]\boldsymbol{\Sigma}_r^{-1}\mathbf{R}\mathbf{M}_r$, $\langle \mathbf{Z}_n\mathbf{Z}_n^\top\rangle_r = \mathbf{M}_r\mathbf{R}^\top\boldsymbol{\Sigma}_r^{-1}\langle \mathbf{Y}_n^r\mathbf{Z}_n^\top\rangle_r + q^c\mathbf{M}_r$, and $\mathbf{H}_n^{(v,r)} = \mathbf{Y}_n^{(v,r)} - \mathbf{R}^{(v)}\mathbf{Z}_n^\top$.

More Experimental Results

Table 3 shows the average rank-one matching accuracy on the PIE data set for poses 22.5° vs. 0°, −22.5° vs. 0°, and −22.5° vs. 22.5°. Table 4 shows the photo-sketch recognition performance on the CUFS data set in the sketch vs. photo settings. As can be seen, the trend and behavior of these results are consistent with those shown in the paper.

Table 3: Average rank-one matching accuracy on the PIE data set (**Best**; Second best).

Pose	22.5° vs. 0°	−22.5° vs. 0°	−22.5° vs. 22.5°
CCA	91.55 ± 3.55	90.38 ± 5.36	72.09 ± 8.39
2DCCA	90.85 ± 5.70	85.36 ± 5.71	54.92 ± 7.83
MCCA1+2	88.51 ± 7.23	88.81 ± 6.07	70.16 ± 10.79
PCCA	**92.98 ± 4.06**	90.82 ± 5.24	74.13 ± 7.14
BMTF	90.11 ± 4.09	88.57 ± 5.02	68.45 ± 10.87
BPCCA	**94.38 ± 3.81**	**92.92 ± 3.73**	**79.67 ± 8.26**



(a) PIE (b) CUHK (c) AR

Figure 2: Log-likelihood of BPCCA over iterations on the PIE, CUHK, and AR data sets.

Convergence Study

We show the log-likelihood of $p(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$ over iterations for BPCCA (with the regularization parameter $\gamma = 0$) on the PIE (22.5° vs. −22.5°), CUHK (photo vs. sketch), and AR (photo vs. sketch) data sets in Figure 2. From the figure, BPCCA does improve the log-likelihood on the whole, although slight drop happens during iterations for the CUHK and AR data sets. The possible cause is that BPCCA does

Table 4: Average rank-one matching accuracy on the CUFS data set (**Best**; Second best).

Matching Type		Sketch vs. Photo			Cropped Sketch vs. Original Photo with $T = 50$			
Training Setting		$T = 25$	$T = 50$	$T = 75$	$m = 2$	$m = 4$	$m = 6$	$m = 8$
CUHK	CCA	40.31 ± 2.13	64.06 ± 4.40	81.50 ± 2.78	72.61 ± 4.39	71.16 ± 3.77	59.06 ± 3.73	30.58 ± 3.15
	2DCCA	74.05 ± 16.73	85.07 ± 17.87	92.65 ± 4.17	58.48 ± 16.34	38.77 ± 10.81	17.97 ± 2.87	7.61 ± 3.42
	MCCA1+2	53.50 ± 5.28	70.80 ± 7.26	81.42 ± 6.00	73.48 ± 5.45	69.20 ± 4.21	49.71 ± 2.96	29.42 ± 3.67
	PCCA	57.42 ± 3.45	80.65 ± 3.17	90.27 ± 2.80	80.43 ± 2.84	78.04 ± 3.90	66.09 ± 3.33	38.70 ± 4.17
	BMTF	95.95 ± 1.16	97.90 ± 0.63	98.50 ± 0.84	77.39 ± 4.71	70.29 ± 6.81	55.36 ± 6.32	31.88 ± 5.17
	BPCCA	$\mathbf{98.47 \pm 0.43}$	$\mathbf{99.13 \pm 0.46}$	$\mathbf{99.03 \pm 0.50}$	$\mathbf{98.99 \pm 0.37}$	$\mathbf{99.20 \pm 0.53}$	$\mathbf{98.62 \pm 1.10}$	$\mathbf{93.62 \pm 8.40}$
AR	CCA	17.35 ± 3.33	29.04 ± 5.16	48.54 ± 5.38	27.53 ± 3.73	27.12 ± 4.65	$\mathbf{21.51 \pm 7.13}$	13.97 ± 4.46
	2DCCA	16.43 ± 5.06	22.19 ± 3.86	32.08 ± 10.68	20.68 ± 3.84	16.58 ± 3.50	11.51 ± 4.88	4.66 ± 1.73
	MCCA1+2	22.14 ± 5.27	35.89 ± 3.34	51.67 ± 4.59	33.42 ± 6.10	30.55 ± 3.60	$\mathbf{22.88 \pm 4.83}$	$\mathbf{14.93 \pm 2.77}$
	PCCA	21.22 ± 4.73	35.89 ± 6.95	54.58 ± 5.54	32.74 ± 5.22	32.33 ± 4.44	$\mathbf{25.21 \pm 5.67}$	$\mathbf{17.81 \pm 2.81}$
	BMTF	28.67 ± 4.50	38.08 ± 4.82	55.21 ± 7.62	16.30 ± 5.34	15.34 ± 3.47	12.33 ± 2.89	8.63 ± 1.45
	BPCCA	$\mathbf{40.31 \pm 4.77}$	$\mathbf{51.64 \pm 4.70}$	$\mathbf{68.54 \pm 5.68}$	$\mathbf{48.36 \pm 3.71}$	$\mathbf{42.88 \pm 6.56}$	$\mathbf{27.40 \pm 5.17}$	$\mathbf{16.16 \pm 4.82}$

not have the monotonicity property of EM, since it maximizes the expectations of $\mathcal{L}_c(\boldsymbol{\theta}_c)$ and $\mathcal{L}_r(\boldsymbol{\theta}_r)$ alternatively instead of the complete-data log-likelihood $\mathcal{L}(\boldsymbol{\theta}_c, \boldsymbol{\theta}_r) = \mathcal{L}_c(\boldsymbol{\theta}_c) + \sum_{n=1}^{N} p(\boldsymbol{x}_n^c | \boldsymbol{y}_n^c) = \mathcal{L}_r(\boldsymbol{\theta}_r) + \sum_{n=1}^{N} p(\boldsymbol{x}_n^r | \boldsymbol{y}_n^r)$. Further investigations are needed for a deeper understanding. Nevertheless, empirically BPCCA is stable and often converges within a relatively small number of iterations.