Learning the Face Prior for Bayesian Face Recognition (Supplementary Materials)

Chaochao Lu and Xiaoou Tang

Department of Information Engineering
The Chinese University of Hong Kong

1 More Experiments

In Section 4.3, we have stated that the maximization of the following log-likelihood easily leads to overfitting,

$$\mathcal{L}(\boldsymbol{\theta}^{\mathcal{K}}) = \sum_{k=1}^{K} \ln \sum_{l=1}^{L} \lambda_l \mathcal{N}(\mathbf{x}^k | \boldsymbol{\mu}_{\mathbf{x}}^l, \boldsymbol{\Sigma}_{\mathbf{x}}^l), \tag{1}$$

which has been analyzed in references [25,40]. Therefore, we propose the leave-set-out (LSO) method to prevent overfitting as follows,

$$\mathcal{L}_{LSO}(\boldsymbol{\theta}^{\mathcal{K}}) = \sum_{l=1}^{L} \sum_{k \in I_l} \ln \sum_{l' \neq l} \lambda_{l'} \mathcal{N}(\mathbf{x}^k | \boldsymbol{\mu}_{\mathbf{x}}^{l'}, \boldsymbol{\Sigma}_{\mathbf{x}}^{l'}). \tag{2}$$

The advantages of the above Equation (2) have been presented in Section 4.3 and 4.5.

Here, we demonstrate that the LSO method provides some protection against overfitting by training on one dataset and then testing on different datasets. For convenience, we train our method with $\mathcal{L}$ and $\mathcal{L}_{LSO}$ on Multi-PIE, PubFig, and WDRef, respectively, and test on View 2 of LFW under the unrestricted protocol. To seek the values of c and L, the View 1 of LFW is used as the validation set. In the experiments below, only LBP feature is extracted in each rectified holistic face as described in Section 5.1. Besides, 10,000 matched pairs and 10,000 mismatched pairs are constructed on each training set during the training procedure. When our method is trained on Multi-PIE, we use the same experiment setting (except that all individuals are used here) as described in Section 5.5, and then the optimal L is estimated using **Method 1**. When our method is trained on PubFig, we also use the same experiment setting (200 individuals and 200 images per individual) as described in Section 5.3. When our method is trained on WDRef, we choose a subset of WDRef with the individuals containing at least 30 images as same as Section 5.7. Using **Method 2**, we can obtain the optimal values of c and L. As shown in Table 1, the performance of our method trained with $\mathcal{L}$ drops much more significantly than $\mathcal{L}_{LSO}$ with

	Multi-PIE → LFW	PubFig → LFW	WDRef → LFW
$\mathcal{L}$	82.5	88.4	91.2
$\mathcal{L}_{LSO}$	**88.6**	**91.1**	**93.3**

Table 1. Accuracies (%) for transfer across datasets

the growing differences between the training dataset and the test dataset. For example, compared with the performance with $\mathcal{L}_{LSO}$, the performance with $\mathcal{L}$ decreases 6.1% when trained on Multi-PIE and tested on LFW, and decreases 2.1% when trained on WDRef and tested on LFW, because LFW is much more different from Multi-PIE than WDRef. Therefore, the LSO method can be used to avoid overfitting.