Learning Canonical Correlations of Paired Tensor Sets Via Tensor-to-Vector Projection

Haiping Lu
Institute for Infocomm Research
Singapore
hplu@ieee.org

Abstract

Canonical correlation analysis (CCA) is a useful technique for measuring relationship between two sets of vector data. For paired tensor data sets, we propose a multilinear CCA (MCCA) method. Unlike existing multilinear variations of CCA, MCCA extracts uncorrelated features under two architectures while maximizing paired correlations. Through a pair of tensor-to-vector projections, one architecture enforces zero-correlation within each set while the other enforces zero-correlation between different pairs of the two sets. We take a successive and iterative approach to solve the problem. Experiments on matching faces of different poses show that MCCA outperforms CCA and 2D-CCA, while using much fewer features. In addition, the fusion of two architectures leads to performance improvement, indicating complementary information.

1 Introduction

Canonical correlation analysis (CCA) [Hotelling, 1936] is a well-known method for analyzing the relations between two sets of vector-valued variables. It assumes that the two data sets are two views of the same set of objects and projects them into low dimensional spaces where each pair is maximally correlated subject to being uncorrelated with other pairs. CCA has various applications such as information retrieval [Hardoon *et al.*, 2004], multi-label learning [Sun *et al.*, 2008; Rai and Daumé III, 2009] and multi-view learning [Chaudhuri *et al.*, 2009; Dhillon *et al.*, 2011].

Many real-world data are multi-dimensional, represented as *tensors* rather than vectors. The number of dimensions is the *order* of a tensor. Each dimension is a *mode*. Second-order tensors include matrix data such as 2-D images. Examples of third-order tensors are video sequences, 3-D images, and web graph mining data organized in three modes of source, destination and text [Kolda and Bader, 2009; Kolda and Sun, 2008].

To deal with multi-dimensional data effectively, researchers have attempted to learn features *directly* from tensors [He *et al.*, 2005]. The motivation is to preserve data structure in feature extraction, obtain more compact representation, and process big data more efficiently, without reshaping tensors into vectors. Many *multilinear subspace learning (MSL)* algorithms [Lu *et al.*, 2011] generalize linear algorithms such as principal component analysis and linear discriminant analysis to second/higher-order [Ye *et al.*, 2004a; 2004b; Yan *et al.*, 2005; Tao *et al.*, 2007; Lu *et al.*, 2008; 2009a; 2009b]. There have also been efforts to generalize CCA to tensor data. To the best of our knowledge, there are two approaches.

One approach analyzes the relationship between two tensor data sets rather than vector data sets. The two-dimensional CCA (2D-CCA) [Lee and Choi, 2007] is the first work along this line to analyze relations between two sets of image data without reshaping into vectors. It was further extended to local 2D-CCA [Wang, 2010], sparse 2D-CCA [Yan *et al.*, 2012], and 3-D CCA (3D-CCA) [Gang *et al.*, 2011]. Under the MSL framework, these CCA extensions are using the *tensor-to-tensor projection (TTP)* [Lu *et al.*, 2011].

Another approach studies the correlations between two data tensors with one or more shared modes because CCA can be viewed as taking two data matrices as input. This idea was first introduced in [Harshman, 2006]. Tensor CCA (TCCA) [Kim and Cipolla, 2009] was proposed later with two architectures for matching (3-D) video volume data of the same dimensions. The single-shared-mode and joint-shared-mode TCCAs can be viewed as *transposed* versions of 2D-CCA and classical CCA, respectively. They apply canonical transformations to the non-shared one/two axes of 3-D data, which can be considered as *partial TTPs* under MSL. Following TCCA, the multiway CCA in [Zhang *et al.*, 2011] uses partial projections to find correlations between second-order and third-order tensors.

However, none of the multilinear extensions above takes an important property of CCA into account, i.e., CCA extracts *uncorrelated features* for different pairs of projections. To close this gap, we propose a *multilinear CCA (MCCA)* algorithm for learning canonical correlations of *paired tensor data sets* with uncorrelated features under *two architectures*. It belongs to the first approach mentioned above.

MCCA uses the *tensor-to-vector projection (TVP)* [Lu *et al.*, 2009a], which will be briefly reviewed in Sec. 2. We follow the CCA derivation in [Anderson, 2003] to successively solve for *elementary multilinear projections (EMPs)* to maximize pairwise correlations while enforcing same-set or cross-set zero-correlation constraint in Sec. 3. Finally, we evaluate the performance on facial image matching in Sec. 4.

To reduce notation complexity, we present our work here only for *second-order tensors*, i.e., matrices, as a special case, although we have developed MCCA for general tensors of any order following [Lu *et al.*, 2011].

2 Preliminary of CCA and TVP

2.1 CCA and zero-correlation constraint

CCA captures the correlations between two sets of vector-valued variables that are assumed to be different representations of the same set of objects [Hotelling, 1936]. For two paired data sets $\mathbf{x}_m \in \mathbb{R}^I$, $\mathbf{y}_m \in \mathbb{R}^J$, $m = 1, ..., M$, $I \neq J$ in general, CCA finds paired projections $\{\mathbf{u}_{x_p}, \mathbf{u}_{y_p}\}$, $\mathbf{u}_{x_p} \in \mathbb{R}^I$, $\mathbf{u}_{y_p} \in \mathbb{R}^J$, $p = 1, ..., P$, such that the canonical variates $\mathbf{w}_p$ and $\mathbf{z}_p$ are maximally correlated. The mth elements of $\mathbf{w}_p$ and $\mathbf{z}_p$ are denoted as w_{p_m} and z_{p_m}, respectively:

$$w_{p_m} = \mathbf{u}'_{x_p}\mathbf{x}_m, \quad z_{p_m} = \mathbf{u}'_{y_p}\mathbf{y}_m, \tag{1}$$

where $\mathbf{u}'$ denotes the transpose of $\mathbf{u}$. In addition, for $p \neq q$ ($p, q = 1, ..., P$), $\mathbf{w}_p$ and $\mathbf{w}_q$, $\mathbf{z}_p$ and $\mathbf{z}_q$, and $\mathbf{w}_p$ and $\mathbf{z}_q$, respectively, are all uncorrelated. Thus, CCA produces uncorrelated features both within each set and between different pairs of the two sets.

The solutions $\{\mathbf{u}_{x_p}, \mathbf{u}_{y_p}\}$ for this CCA problem can be obtained as eigenvectors of two related eigenvalue problems. However, a straightforward generalization of these results to tensors [Lee and Choi, 2007; Gang *et al.*, 2011] does not give uncorrelated features while maximizing the paired correlations (to be shown in Figs. 3(b) and 3(c)). Therefore, we derive MCCA following the more principled approach of solving CCA in [Anderson, 2003], where CCA projection pairs are obtained successively through enforcing the zero-correlation constraints.

2.2 TVP: tensor space to vector subspace

To extract uncorrelated features from tensor data directly, we use the TVP in [Lu *et al.*, 2009a]. The TVP of a matrix $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2}$ to a vector $\mathbf{r} \in \mathbb{R}^P$ consists of P EMPs $\{\mathbf{u}_p, \mathbf{v}_p\}$, $\mathbf{u}_p \in \mathbb{R}^{I_1}$, $\mathbf{v}_p \in \mathbb{R}^{I_2}$, $p = 1, ..., P$. It is a mapping from the original tensor space $\mathbb{R}^{I_1} \bigotimes \mathbb{R}^{I_2}$ into a vector subspace $\mathbb{R}^P$. The pth EMP $(\mathbf{u}_p, \mathbf{v}_p)$ projects $\mathbf{X}$ to a scalar r_p, the pth element of $\mathbf{r}$, as:

$$r_p = \mathbf{u}'_p\mathbf{X}\mathbf{v}_p. \tag{2}$$

Define two matrices $\mathbf{U} = [\mathbf{u}_1, ..., \mathbf{u}_P] \in \mathbb{R}^{I_1 \times P}$ and $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_P] \in \mathbb{R}^{I_2 \times P}$, then we have

$$\mathbf{r} = diag\left(\mathbf{U}'\mathbf{X}\mathbf{V}\right), \tag{3}$$

where $diag(\cdot)$ denotes the main diagonal of a matrix.

3 Multilinear CCA

This section proposes MCCA for analyzing the correlations between two matrix data sets by first formulating the problem, and then adopting the successive maximization approach [Anderson, 2003] and alternating projection method [Lu *et al.*, 2011] to solve it.

3.1 Problem formulation through TVP

In (second-order) MCCA, we consider two matrix data sets with M samples each: $\left\{\mathbf{X}_m \in \mathbb{R}^{I_1 \times I_2}\right\}$ and $\left\{\mathbf{Y}_m \in \mathbb{R}^{J_1 \times J_2}\right\}$, $m = 1, ..., M$. I_1 and I_2 may not equal to J_1 and J_2, respectively. We assume that their means have been subtracted so that they have *zero-mean*, i.e., $\sum_{m=1}^{M}\mathbf{X}_m = \mathbf{0}$ and $\sum_{m=1}^{M}\mathbf{Y}_m = \mathbf{0}$. We are interested in paired projections $\{(\mathbf{u}_{x_p}, \mathbf{v}_{x_p}), (\mathbf{u}_{y_p}, \mathbf{v}_{y_p})\}$, $\mathbf{u}_{x_p} \in \mathbb{R}^{I_1}$, $\mathbf{v}_{x_p} \in \mathbb{R}^{I_2}$, $\mathbf{u}_{y_p} \in \mathbb{R}^{J_1}$, $\mathbf{v}_{y_p} \in \mathbb{R}^{J_2}$, $p = 1, ..., P$, for *dimensionality reduction* to vector spaces to reveal correlations between them. The P projection pairs form matrices $\mathbf{U}_x = [\mathbf{u}_{x_1}, ..., \mathbf{u}_{x_P}]$, $\mathbf{V}_x = [\mathbf{v}_{x_1}, ..., \mathbf{v}_{x_P}]$, $\mathbf{U}_y = [\mathbf{u}_{y_1}, ..., \mathbf{u}_{y_P}]$, and $\mathbf{V}_y = [\mathbf{v}_{y_1}, ..., \mathbf{v}_{y_P}]$. The matrix sets $\{\mathbf{X}_m\}$ and $\{\mathbf{Y}_m\}$ are projected to $\{\mathbf{r}_m \in \mathbb{R}^P\}$ and $\{\mathbf{s}_m \in \mathbb{R}^P\}$, respectively, through TVP (3) as

$$\mathbf{r}_m = diag\left(\mathbf{U}'_x\mathbf{X}_m\mathbf{V}_x\right), \quad \mathbf{s}_m = diag\left(\mathbf{U}'_y\mathbf{Y}_m\mathbf{V}_y\right). \tag{4}$$

We can write the projection above according to EMP (2) as

$$r_{m_p} = \mathbf{u}'_{x_p}\mathbf{X}_m\mathbf{v}_{x_p}, \quad s_{m_p} = \mathbf{u}'_{y_p}\mathbf{Y}_m\mathbf{v}_{y_p}, \tag{5}$$

where r_{m_p} and s_{m_p} are the pth elements of $\mathbf{r}_m$ and $\mathbf{s}_m$, respectively. We then define two *coordinate vectors* $\mathbf{w}_p \in \mathbb{R}^M$ and $\mathbf{z}_p \in \mathbb{R}^M$ for the pth EMP, where their pth elements $w_{p_m} = r_{m_p}$ and $z_{p_m} = s_{m_p}$. Denote $\mathbf{R} = [\mathbf{r}_1, ..., \mathbf{r}_M]$, $\mathbf{S} = [\mathbf{s}_1, ..., \mathbf{s}_M]$, $\mathbf{W} = [\mathbf{w}_1, ..., \mathbf{w}_P]$, and $\mathbf{Z} = [\mathbf{z}_1, ..., \mathbf{z}_P]$. We have $\mathbf{W} = \mathbf{R}'$ and $\mathbf{Z} = \mathbf{S}'$. Here, $\mathbf{w}_p$ and $\mathbf{z}_p$ are analogous to the *canonical variates* in CCA [Anderson, 2003].

We propose MCCA with two architectures. **Architecture I:** $\mathbf{w}_p$ and $\mathbf{z}_p$ are maximally correlated, while for $q \neq p$ ($p, q = 1, ..., P$), $\mathbf{w}_q$ and $\mathbf{w}_p$, and $\mathbf{z}_q$ and $\mathbf{z}_p$, respectively, are uncorrelated. **Architecture II:** $\mathbf{w}_p$ and $\mathbf{z}_p$ are maximally correlated, while for $q \neq p$ ($p, q = 1, ..., P$), $\mathbf{w}_p$ and $\mathbf{z}_q$ are uncorrelated. Figure 1 is a schematic diagram showing MCCA and its two different architectures.

We extend the sample-based estimation of canonical correlations and variates in CCA to the multilinear case for matrix data using Architecture I.

Definition 1. ***Canonical correlations and variates for matrix data (Architecture I).*** *For two matrix data sets with M samples each $\{\mathbf{X}_m \in \mathbb{R}^{I_1 \times I_2}\}$ and $\{\mathbf{Y}_m \in \mathbb{R}^{J_1 \times J_2}\}$, $m = 1, ..., M$, the pth pair of canonical variates is the pair $\{w_{p_m} = \mathbf{u}'_{x_p}\mathbf{X}_m\mathbf{v}_{x_p}\}$ and $\{z_{p_m} = \mathbf{u}'_{y_p}\mathbf{Y}_m\mathbf{v}_{y_p}\}$, where $\mathbf{w}_p$ and $\mathbf{z}_p$ have maximum correlation ρ_p while for $q \neq p$ ($p, q = 1, ..., P$), $\mathbf{w}_q$ and $\mathbf{w}_p$, and $\mathbf{z}_q$ and $\mathbf{z}_p$, respectively, are uncorrelated. The correlation ρ_p is the pth canonical correlation.*

An analogous definition with Architecture II follows by changing the constraint accordingly.

3.2 Successive derivation of MCCA algorithm

Here, we derive MCCA with *Architecture I* (MCCA1). MCCA with *Architecture II* (MCCA2) differs from MCCA1 only in the constraints enforced so it can be obtained in an analogous way by swapping $\mathbf{w}_q$ and $\mathbf{z}_q$ ($q = 1, ..., p-1$) below. Following the successive derivation of CCA in [Anderson, 2003], we consider the canonical correlations one by

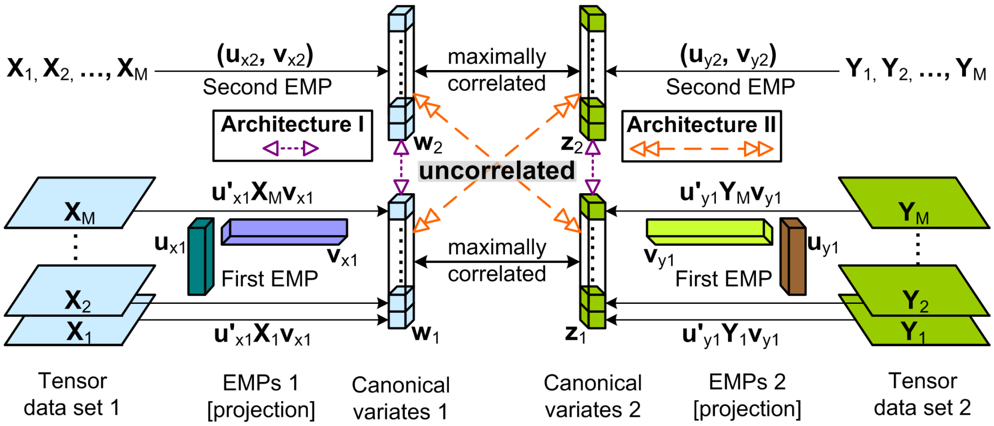

Figure 1: Schematic of multilinear CCA for paired tensor data sets with two architectures. EMP stands for elementary multilinear projection, which projects a tensor to a scalar. *Matrices* are viewed as *second-order tensors*.

one. The MCCA objective function for the pth EMP pair is

$$\{(\mathbf{u}_{x_p}, \mathbf{v}_{x_p}), (\mathbf{u}_{y_p}, \mathbf{v}_{y_p})\} = \arg\max \rho_p, \texttt{subject to}$$
$$\mathbf{w}'_p \mathbf{w}_q = 0, \mathbf{z}'_p \mathbf{z}_q = 0, \texttt{ for } p \neq q,\ p, q = 1, ..., P, \tag{6}$$

where the *sample* Pearson correlation coefficient ρ_p between $\mathbf{w}_p$ and $\mathbf{z}_p$ is given by[1]:

$$\rho_p = \frac{\mathbf{w}'_p \mathbf{z}_p}{\|\mathbf{w}_p\| \|\mathbf{z}_p\|}. \tag{7}$$

The P EMP pairs in the MCCA problem are determined one pair at a time in P steps, with the pth step obtaining the pth EMP pair:

Step 1**:** Determine $\{(\mathbf{u}_{x_1}, \mathbf{v}_{x_1}), (\mathbf{u}_{y_1}, \mathbf{v}_{y_1})\}$ that maximizes ρ_1 without any constraint.

Step $p(= 2, ..., P)$**:** Determine $\{(\mathbf{u}_{x_p}, \mathbf{v}_{x_p}), (\mathbf{u}_{y_p}, \mathbf{v}_{y_p})\}$ that maximizes ρ_p subject to the constraint that $\mathbf{w}'_p \mathbf{w}_q = \mathbf{z}'_p \mathbf{z}_q = 0$ for $q = 1, ..., p-1$.

To solve for the pth EMP pair, we need to determine four projection vectors, $\mathbf{u}_{x_p}, \mathbf{v}_{x_p}, \mathbf{u}_{y_p}$, and $\mathbf{v}_{y_p}$. As their simultaneous determination is difficult, we follow the *alternating projection method* [Lu *et al.*, 2011] derived from *alternating least squares* [Harshman, 1970]. To determine each EMP pair, we estimate $\mathbf{u}_{x_p}$ and $\mathbf{u}_{y_p}$ conditioned on $\mathbf{v}_{x_p}$ and $\mathbf{v}_{y_p}$ first, then we estimate $\mathbf{v}_{x_p}$ and $\mathbf{v}_{y_p}$ conditioned on $\mathbf{u}_{x_p}$ and $\mathbf{u}_{y_p}$.

Conditional subproblem: To estimate $\mathbf{u}_{x_p}$ and $\mathbf{u}_{y_p}$ conditioned on $\mathbf{v}_{x_p}$ and $\mathbf{v}_{y_p}$, we assume that $\mathbf{v}_{x_p}$ and $\mathbf{v}_{y_p}$ are given and they can project input samples to obtain vectors

$$\tilde{\mathbf{r}}_{m_p} = \mathbf{X}_m \mathbf{v}_{x_p} \in \mathbb{R}^{I_1}, \quad \tilde{\mathbf{s}}_{m_p} = \mathbf{Y}_m \mathbf{v}_{y_p} \in \mathbb{R}^{J_1}. \tag{8}$$

The conditional subproblem then becomes to determine $\mathbf{u}_{x_p}$ and $\mathbf{u}_{y_p}$ that project the samples $\{\tilde{\mathbf{r}}_{m_p}\}$ and $\{\tilde{\mathbf{s}}_{m_p}\}$ ($m = 1, ..., M$) to maximize their correlation subject to the *zero-correlation* constraint, which is a CCA problem with input samples $\{\tilde{\mathbf{r}}_{m_p}\}$ and $\{\tilde{\mathbf{s}}_{m_p}\}$.

The corresponding *maximum likelihood estimator* of the covariance matrices are defined as [Anderson, 2003]

$$\tilde{\mathbf{\Sigma}}_{\mathbf{rr}_p} = \frac{1}{M}\sum_{m=1}^{M} \tilde{\mathbf{r}}_{m_p} \tilde{\mathbf{r}}'_{m_p}, \quad \tilde{\mathbf{\Sigma}}_{\mathbf{ss}_p} = \frac{1}{M}\sum_{m=1}^{M} \tilde{\mathbf{s}}_{m_p} \tilde{\mathbf{s}}'_{m_p}, \tag{9}$$

$$\tilde{\mathbf{\Sigma}}_{\mathbf{rs}_p} = \frac{1}{M}\sum_{m=1}^{M} \tilde{\mathbf{r}}_{m_p} \tilde{\mathbf{s}}'_{m_p}, \quad \tilde{\mathbf{\Sigma}}_{\mathbf{sr}_p} = \frac{1}{M}\sum_{m=1}^{M} \tilde{\mathbf{s}}_{m_p} \tilde{\mathbf{r}}'_{m_p}, \tag{10}$$

where $\tilde{\mathbf{\Sigma}}_{\mathbf{rs}_p} = \tilde{\mathbf{\Sigma}}'_{\mathbf{sr}_p}$.

For $p = 1$, we solve for $\mathbf{u}_{x_1}$ and $\mathbf{u}_{y_1}$ that maximize the correlation between the projections of $\{\tilde{\mathbf{r}}_{m_p}\}$ and $\{\tilde{\mathbf{s}}_{m_p}\}$ without any constraint. We get $\mathbf{u}_{x_1}$ and $\mathbf{u}_{y_1}$ as the leading eigenvectors of the following eigenvalue problems as in CCA [Hardoon *et al.*, 2004; Dhillon *et al.*, 2011]:

$$\tilde{\mathbf{\Sigma}}_{\mathbf{rr}_p}^{-1} \tilde{\mathbf{\Sigma}}_{\mathbf{rs}_p} \tilde{\mathbf{\Sigma}}_{\mathbf{ss}_p}^{-1} \tilde{\mathbf{\Sigma}}_{\mathbf{sr}_p} \mathbf{u}_{x_1} = \lambda \mathbf{u}_{x_1}, \tag{11}$$

$$\tilde{\mathbf{\Sigma}}_{\mathbf{ss}_p}^{-1} \tilde{\mathbf{\Sigma}}_{\mathbf{sr}_p} \tilde{\mathbf{\Sigma}}_{\mathbf{rr}_p}^{-1} \tilde{\mathbf{\Sigma}}_{\mathbf{rs}_p} \mathbf{u}_{y_1} = \lambda \mathbf{u}_{y_1}, \tag{12}$$

where we assume that $\tilde{\mathbf{\Sigma}}_{\mathbf{ss}_p}$ and $\tilde{\mathbf{\Sigma}}_{\mathbf{rr}_p}$ are *nonsingular* throughout this paper.

Constrained optimization for $p > 1$**:** Next, we determine the pth ($p > 1$) EMP pair given the first $(p-1)$ EMP pairs by maximizing the correlation ρ_p, subject to the constraint that features projected by the pth EMP pair are *uncorrelated* with those projected by the first $(p-1)$ EMP pairs. Due to the fundamental difference between linear and multilinear projections, the remaining eigenvectors of (11) and (12), which are the solutions of CCA for $p > 1$, are not the solutions for our MCCA problem. Instead, we have to solve a ***new constrained optimization problem* in the *multilinear setting***.

Let $\tilde{\mathbf{R}}_p \in \mathbb{R}^{I_1 \times M}$ and $\tilde{\mathbf{S}}_p \in \mathbb{R}^{J_1 \times M}$ be matrices with $\tilde{\mathbf{r}}_{m_p}$ and $\tilde{\mathbf{s}}_{m_p}$ as their mth columns, respectively, i.e.,

$$\tilde{\mathbf{R}}_p = \left[\tilde{\mathbf{r}}_{1_p}, \tilde{\mathbf{r}}_{2_p}, ..., \tilde{\mathbf{r}}_{M_p}\right], \quad \tilde{\mathbf{S}}_p = \left[\tilde{\mathbf{s}}_{1_p}, \tilde{\mathbf{s}}_{2_p}, ..., \tilde{\mathbf{s}}_{M_p}\right], \tag{13}$$

[1]Note that the means of $\mathbf{w}_p$ and $\mathbf{z}_p$ are both zero since the input matrix sets are assumed to be *zero-mean*.

then the pth coordinate vectors are $\mathbf{w}_p = \tilde{\mathbf{R}}'_p \mathbf{u}_{x_p}$ and $\mathbf{z}_p = \tilde{\mathbf{S}}'_p \mathbf{u}_{y_p}$. The constraint that $\mathbf{w}_p$ and $\mathbf{z}_p$ are uncorrelated with $\{\mathbf{w}_q\}$ and $\{\mathbf{z}_q\}$, respectively, can be written as

$$\mathbf{u}'_{x_p}\tilde{\mathbf{R}}_p\mathbf{w}_q = 0, \mathbf{u}'_{y_p}\tilde{\mathbf{S}}_p\mathbf{z}_q = 0, q = 1, ..., p-1. \quad (14)$$

The *coordinate matrices* are formed as

$$\mathbf{W}_{p-1} = [\mathbf{w}_1 \quad \mathbf{w}_2 \quad ...\mathbf{w}_{p-1}] \in \mathbb{R}^{M\times(p-1)}, \quad (15)$$

$$\mathbf{Z}_{p-1} = [\mathbf{z}_1 \quad \mathbf{z}_2 \quad ...\mathbf{z}_{p-1}] \in \mathbb{R}^{M\times(p-1)}. \quad (16)$$

Thus, we determine $\mathbf{u}_{x_p}$ and $\mathbf{u}_{y_p}$ ($p > 1$) by solving the following constrained optimization problem:

$$\{\mathbf{u}_{x_p}, \mathbf{u}_{y_p}\} = \arg\max \tilde{\rho}_p, \quad \texttt{subject to} \quad (17)$$
$$\mathbf{u}'_{x_p}\tilde{\mathbf{R}}_p\mathbf{w}_q = 0, \mathbf{u}'_{y_p}\tilde{\mathbf{S}}_p\mathbf{z}_q = 0, q = 1, ..., p-1,$$

where the pth (sample) canonical correlation

$$\tilde{\rho}_p = \frac{\mathbf{u}'_{x_p}\tilde{\boldsymbol{\Sigma}}_{\mathbf{rs}_p}\mathbf{u}_{y_p}}{\sqrt{\mathbf{u}'_{x_p}\tilde{\boldsymbol{\Sigma}}_{\mathbf{rr}_p}\mathbf{u}_{x_p}\mathbf{u}'_{y_p}\tilde{\boldsymbol{\Sigma}}_{\mathbf{ss}_p}\mathbf{u}_{y_p}}}. \quad (18)$$

The solution is given by the following theorem:

Theorem 1. *For nonsingular $\tilde{\boldsymbol{\Sigma}}_{\mathbf{rr}_p}$ and $\tilde{\boldsymbol{\Sigma}}_{\mathbf{ss}_p}$, the solutions for $\mathbf{u}_{x_p}$ and $\mathbf{u}_{y_p}$ in the problem (17) are the eigenvectors corresponding to the largest eigenvalue λ of the following eigenvalue problems:*

$$\tilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{rr}_p}\boldsymbol{\Phi}_p\tilde{\boldsymbol{\Sigma}}_{\mathbf{rs}_p}\tilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{ss}_p}\boldsymbol{\Psi}_p\tilde{\boldsymbol{\Sigma}}_{\mathbf{sr}_p}\mathbf{u}_{x_p} = \lambda\mathbf{u}_{x_p}, \quad (19)$$

$$\tilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{ss}_p}\boldsymbol{\Psi}_p\tilde{\boldsymbol{\Sigma}}_{\mathbf{sr}_p}\tilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{rr}_p}\boldsymbol{\Phi}_p\tilde{\boldsymbol{\Sigma}}_{\mathbf{rs}_p}\mathbf{u}_{y_p} = \lambda\mathbf{u}_{y_p}. \quad (20)$$

where

$$\boldsymbol{\Phi}_p = \mathbf{I} - \tilde{\mathbf{R}}_p\mathbf{W}_{p-1}\boldsymbol{\Theta}^{-1}_p\mathbf{W}'_{p-1}\tilde{\mathbf{R}}'_p\tilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{rr}_p}, \quad (21)$$

$$\boldsymbol{\Psi}_p = \mathbf{I} - \tilde{\mathbf{S}}_p\mathbf{Z}_{p-1}\boldsymbol{\Omega}^{-1}_p\mathbf{Z}'_{p-1}\tilde{\mathbf{S}}'_p\tilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{ss}_p}, \quad (22)$$

$$\boldsymbol{\Theta}_p = \mathbf{W}'_{p-1}\tilde{\mathbf{R}}'_p\tilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{rr}_p}\tilde{\mathbf{R}}_p\mathbf{W}_{p-1}, \quad (23)$$

$$\boldsymbol{\Omega}_p = \mathbf{Z}'_{p-1}\tilde{\mathbf{S}}'_p\tilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{ss}_p}\tilde{\mathbf{S}}_p\mathbf{Z}_{p-1}, \quad (24)$$

and $\mathbf{I}$ is an identity matrix.

Proof. For nonsingular $\tilde{\boldsymbol{\Sigma}}_{\mathbf{rr}_p}$ and $\tilde{\boldsymbol{\Sigma}}_{\mathbf{ss}_p}$, any $\mathbf{u}_{x_p}$ and $\mathbf{u}_{y_p}$ can be normalized such that $\mathbf{u}'_{x_p}\tilde{\boldsymbol{\Sigma}}_{\mathbf{rr}_p}\mathbf{u}_{x_p} = \mathbf{u}'_{y_p}\tilde{\boldsymbol{\Sigma}}_{\mathbf{ss}_p}\mathbf{u}_{y_p} = 1$ and the ratio $\tilde{\rho}_p$ keeps unchanged. Therefore, maximizing $\tilde{\rho}_p$ is equivalent to maximizing $\mathbf{u}'_{x_p}\tilde{\boldsymbol{\Sigma}}_{\mathbf{rs}_p}\mathbf{u}_{y_p}$ with these normalization constraints. Lagrange multipliers can be used to transform the problem (17) to the following to include all the constraints:

$$\varphi_p = \mathbf{u}'_{x_p}\tilde{\boldsymbol{\Sigma}}_{\mathbf{rs}_p}\mathbf{u}_{y_p} - \frac{\phi}{2}\left(\mathbf{u}'_{x_p}\tilde{\boldsymbol{\Sigma}}_{\mathbf{rr}_p}\mathbf{u}_{x_p} - 1\right) - \frac{\psi}{2}\left(\mathbf{u}'_{y_p}\tilde{\boldsymbol{\Sigma}}_{\mathbf{ss}_p}\mathbf{u}_{y_p} - 1\right) - \sum_{q=1}^{p-1}\theta_q\mathbf{u}'_{x_p}\tilde{\mathbf{R}}_p\mathbf{w}_q - \sum_{q=1}^{p-1}\omega_q\mathbf{u}'_{y_p}\tilde{\mathbf{S}}_p\mathbf{z}_q, \quad (25)$$

where ϕ, ψ, $\{\theta_q\}$, and $\{\omega_q\}$ are Lagrange multipliers. Let $\boldsymbol{\theta}_{p-1} = [\theta_1 \ \theta_2 \ ... \ \theta_{p-1}]'$, $\boldsymbol{\omega}_{p-1} = [\omega_1 \ \omega_2 \ ... \ \omega_{p-1}]'$, then we have the two summations in (25) as $\sum_{q=1}^{p-1}\theta_q\tilde{\mathbf{R}}_p\mathbf{w}_q = \tilde{\mathbf{R}}_p\mathbf{W}_{p-1}\boldsymbol{\theta}_{p-1}$ and $\sum_{q=1}^{p-1}\omega_q\tilde{\mathbf{S}}_p\mathbf{z}_q = \tilde{\mathbf{S}}_p\mathbf{Z}_{p-1}\boldsymbol{\omega}_{p-1}$.

The vectors of partial derivatives of φ_p with respect to the elements of $\mathbf{u}_{x_p}$ and $\mathbf{u}_{y_p}$ are set equal to zeros, giving

$$\frac{\partial\varphi_p}{\partial\mathbf{u}_{x_p}} = \tilde{\boldsymbol{\Sigma}}_{\mathbf{rs}_p}\mathbf{u}_{y_p} - \phi\tilde{\boldsymbol{\Sigma}}_{\mathbf{rr}_p}\mathbf{u}_{x_p} - \tilde{\mathbf{R}}_p\mathbf{W}_{p-1}\boldsymbol{\theta}_{p-1} = \mathbf{0}, \quad (26)$$

$$\frac{\partial\varphi_p}{\partial\mathbf{u}_{y_p}} = \tilde{\boldsymbol{\Sigma}}_{\mathbf{sr}_p}\mathbf{u}_{x_p} - \psi\tilde{\boldsymbol{\Sigma}}_{\mathbf{ss}_p}\mathbf{u}_{y_p} - \tilde{\mathbf{S}}_p\mathbf{Z}_{p-1}\boldsymbol{\omega}_{p-1} = \mathbf{0}. \quad (27)$$

Multiplication of (26) on the left by $\mathbf{u}'_{x_p}$ and (27) on the left by $\mathbf{u}'_{y_p}$ results in

$$\mathbf{u}'_{x_p}\tilde{\boldsymbol{\Sigma}}_{\mathbf{rs}_p}\mathbf{u}_{y_p} - \phi\mathbf{u}'_{x_p}\tilde{\boldsymbol{\Sigma}}_{\mathbf{rr}_p}\mathbf{u}_{x_p} = 0, \quad (28)$$

$$\mathbf{u}'_{y_p}\tilde{\boldsymbol{\Sigma}}'_{\mathbf{rs}_p}\mathbf{u}_{x_p} - \psi\mathbf{u}'_{y_p}\tilde{\boldsymbol{\Sigma}}_{\mathbf{ss}_p}\mathbf{u}_{y_p} = 0. \quad (29)$$

Since $\mathbf{u}'_{x_p}\tilde{\boldsymbol{\Sigma}}_{\mathbf{rr}_p}\mathbf{u}_{x_p} = 1$ and $\mathbf{u}'_{y_p}\tilde{\boldsymbol{\Sigma}}_{\mathbf{ss}_p}\mathbf{u}_{y_p} = 1$, this shows that $\phi = \psi = \mathbf{u}'_{x_p}\tilde{\boldsymbol{\Sigma}}_{\mathbf{rs}_p}\mathbf{u}_{y_p}$.

Next, we have the following using the definitions in (23) and (24) through multiplication of (26) on the left by $\mathbf{W}'_{p-1}\tilde{\mathbf{R}}'_p\tilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{rr}_p}$ and (27) on the left by $\mathbf{Z}'_{p-1}\tilde{\mathbf{S}}'_p\tilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{ss}_p}$:

$$\mathbf{W}'_{p-1}\tilde{\mathbf{R}}'_p\tilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{rr}_p}\tilde{\boldsymbol{\Sigma}}_{\mathbf{rs}_p}\mathbf{u}_{y_p} - \boldsymbol{\Theta}_p\boldsymbol{\theta}_{p-1} = \mathbf{0}, \quad (30)$$

$$\mathbf{Z}'_{p-1}\tilde{\mathbf{S}}'_p\tilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{ss}_p}\tilde{\boldsymbol{\Sigma}}_{\mathbf{sr}_p}\mathbf{u}_{x_p} - \boldsymbol{\Omega}_p\boldsymbol{\omega}_{p-1} = \mathbf{0}. \quad (31)$$

Solving for $\boldsymbol{\theta}_{p-1}$ from (30) and $\boldsymbol{\omega}_{p-1}$ from (31), and substituting them into (26) and (27), we have the following based on earlier result $\phi = \psi$ and the definitions in (21) and (22):

$$\boldsymbol{\Phi}_p\tilde{\boldsymbol{\Sigma}}_{\mathbf{rs}_p}\mathbf{u}_{y_p} = \phi\tilde{\boldsymbol{\Sigma}}_{\mathbf{rr}_p}\mathbf{u}_{x_p}, \ \boldsymbol{\Psi}_p\tilde{\boldsymbol{\Sigma}}_{\mathbf{sr}_p}\mathbf{u}_{x_p} = \phi\tilde{\boldsymbol{\Sigma}}_{\mathbf{ss}_p}\mathbf{u}_{y_p}. \quad (32)$$

Let $\lambda = \phi^2$ (then $\phi = \sqrt{\lambda}$), we have two eigenvalue problems in (19) and (20) from (32). Since $\phi(= \sqrt{\lambda})$ is the criterion to be maximized, the maximization is achieved by setting $\mathbf{u}_{x_p}$ and $\mathbf{u}_{y_p}$ to be the eigenvectors corresponding to the largest eigenvalue of (19) and (20), respectively. □

By setting $\boldsymbol{\Phi}_1 = \mathbf{I}$, $\boldsymbol{\Psi}_1 = \mathbf{I}$, we can write (11) and (12) in the form of (19) and (20) as well. In this way, (19) and (20) give unified solutions for $\{(\mathbf{u}_{x_p}, \mathbf{v}_{x_p}), (\mathbf{u}_{y_p}, \mathbf{v}_{y_p})\}$, $p = 1, ..., P$. They are unique only up to sign [Anderson, 2003].

Similarly, to estimate $\mathbf{v}_{x_p}$ and $\mathbf{v}_{y_p}$ conditioned on $\mathbf{u}_{x_p}$ and $\mathbf{u}_{y_p}$, we project input samples to obtain

$$\tilde{\mathbf{r}}_{m_p} = \mathbf{X}'_m\mathbf{u}_{x_p} \in \mathbb{R}^{I_2}, \quad \tilde{\mathbf{s}}_{m_p} = \mathbf{Y}'_m\mathbf{u}_{y_p} \in \mathbb{R}^{J_2}. \quad (33)$$

This conditional subproblem can be solved similarly by following the derivation from (9) to (24). Algorithm 1 summarizes the MCCA algorithm for Architecture I[2]. For simplicity and reproducibility, we adopt the *uniform initialization* scheme in [Lu *et al.*, 2009b] where each vector is initialized to all ones vector $\mathbf{1}$ and then normalized to unit vector. The iteration is terminated by setting a maximum iteration number K for computational efficiency.

[2]The MATLAB code for MCCA will be made available at: http://www.mathworks.com/matlabcentral/fileexchange/authors/80417

Algorithm 1 Multilinear CCA for matrix sets (Arch. I)

1: **Input:** Two *zero-mean* matrix sets: $\{\mathbf{X}_m \in \mathbb{R}^{I_1 \times I_2}\}$ and $\{\mathbf{Y}_m \in \mathbb{R}^{J_1 \times J_2}\}$, $m = 1, ..., M$, the subspace dimension P, and the maximum number of iterations K.
2: **for** $p = 1$ **to** P **do**
3:    Initialize $\mathbf{v}_{x_{p(0)}} = \mathbf{1}/\parallel \mathbf{1} \parallel$, $\mathbf{v}_{y_{p(0)}} = \mathbf{1}/\parallel \mathbf{1} \parallel$.
4:    **for** $k = 1$ **to** K **do**
5:       **Mode 1:**
6:       Calculate $\{\tilde{\mathbf{r}}_{m_p}\}$ and $\{\tilde{\mathbf{s}}_{m_p}\}$ according to (8) with $\mathbf{v}_{x_{p(k-1)}}$ and $\mathbf{v}_{y_{p(k-1)}}$ for $m = 1, ..., M$.
7:       Calculate $\tilde{\mathbf{\Sigma}}_{\mathbf{rr}_p}$, $\tilde{\mathbf{\Sigma}}_{\mathbf{ss}_p}$, $\tilde{\mathbf{\Sigma}}_{\mathbf{rs}_p}$ and $\tilde{\mathbf{\Sigma}}_{\mathbf{sr}_p}$ from (9) and (10).
8:       Form $\tilde{\mathbf{R}}_p$ and $\tilde{\mathbf{S}}_p$ according to (13).
9:       Calculate $\mathbf{\Theta}_p$, $\mathbf{\Omega}_p$, $\mathbf{\Phi}_p$ and $\mathbf{\Psi}_p$ according to (23), (24), (21) and (22), respectively.
10:       Set $\mathbf{u}_{x_{p(k)}}$ and $\mathbf{u}_{y_{p(k)}}$ to be the eigenvectors of (19) and (20), respectively, associated with the largest eigenvalue.
11:       **Mode 2:**
12:       Calculate $\{\tilde{\mathbf{r}}_{m_p}\}$ and $\{\tilde{\mathbf{s}}_{m_p}\}$ according to (33) with $\mathbf{u}_{x_{p(k)}}$ and $\mathbf{u}_{y_{p(k)}}$ for $m = 1, ..., M$.
13:       Repeat line 7-10, replacing $\mathbf{u}$ with $\mathbf{v}$ in line 10.
14:    **end for**
15:    Set $\mathbf{u}_{x_p} = \mathbf{u}_{x_{p(K)}}$, $\mathbf{u}_{y_p} = \mathbf{u}_{y_{p(K)}}$, $\mathbf{v}_{x_p} = \mathbf{v}_{x_{p(K)}}$, $\mathbf{v}_{y_p} = \mathbf{v}_{y_{p(K)}}$.
16:    Calculate ρ_p and form $\mathbf{W}_{p-1}$ and $\mathbf{Z}_{p-1}$ according to (15) and (16), respectively.
17: **end for**
18: **Output:** $\{(\mathbf{u}_{x_p}, \mathbf{v}_{x_p}), (\mathbf{u}_{y_p}, \mathbf{v}_{y_p}), \rho_p\}$, $p = 1, ..., P$.

3.3 Discussion

Convergence: In a successive optimization step p, each iteration maximizes the correlation between the pth pair of canonical variates $\mathbf{w}_p$ and $\mathbf{z}_p$ subject to zero-correlation constraint. Thus, the captured canonical correlation ρ_p is expected to increase monotonically. However, due to its iterative nature, MCCA can only converge to a local maximum, giving a sub-optimal solution. We will show empirical results on convergence performance in the next section.

MCCA vs. related works: To the best of our knowledge, MCCA is the first multilinear extension of CCA producing same-set/cross-set *uncorrelated features*. It is also the first CCA extension using TVP, which projects a tensor to a *vector*, making constraint enforcement easier. In contrast, 2D-CCA [Lee and Choi, 2007] and 3DCCA [Gang *et al.*, 2011] use TTP to project a tensor to another *tensor without enforcing the zero-correlation constraint*, resulting in features correlated in the same-set and cross-set, as to be shown in Figs. 3(b) and 3(c). On the other hand, TCCA [Kim and Cipolla, 2009] matches two tensors of the *same dimension*. TCCA in the joint-shared-mode requires vectorization of two modes, which is equivalent to classical CCA for vector sets. TCCA in the single-shared-mode solves for transformations in two modes, which is equivalent to 2D-CCA on paired sets consisting of slices of the input tensors.

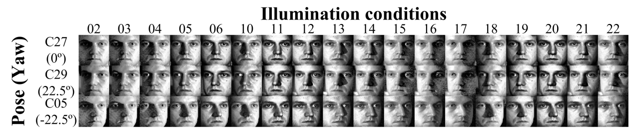

Figure 2: Examples from the selected PIE subset: 54 faces of a subject (3 poses×18 illuminations).

4 Experimental Evaluation

This section evaluates MCCA on learning mappings between facial images of different poses as in [Lee and Choi, 2007].

Data: We selected faces from the PIE database [Sim *et al.*, 2003] to form paired tensor sets of different poses. Three poses C27, C29 and C05 were selected, corresponding to Yaw angle of 0, 22.5 and -22.5 degrees, as shown in Fig. 2. This subset includes faces from 68 subjects captured under 18 illumination conditions (02 to 06 and 10 to 22), where 07 to 09 were excluded due to missing faces. Thus, we have $68 \times 3 \times 18 = 3,672$ faces for this study. All face images were cropped, aligned with annotated eye coordinates, and normalized to 32×32 pixels with 8-bit depth.

Algorithm settings: We denote MCCA with Architectures I and II as *MCCA1* and *MCCA2*, respectively. We also examine their *fusion* to see whether they contain complementary information using *score level fusion* with a simple *sum rule*, denoted as ***MCCA1+2***. They are compared against CCA[3] and 2D-CCA [Lee and Choi, 2007]. We set the maximum number of iterations to ten for 2D-CCA and MCCA. We initialize the projection matrices in 2D-CCA to identity matrices, which gives better performance than random initialization in our study. We extract the maximum number of features for each method (32 for MCCA, 1024 for 2D-CCA and CCA). The extracted features are sorted according to the correlations captured in descending order. Face images are input directly as 32×32 matrices to 2D-CCA and MCCA. For CCA, they are reshaped to 1024×1 vectors as input.

Convergence study: Figure 3(a) depicts an example of the correlations captured over up to 20 iterations during learning for the first five canonical variates of MCCA1. Typically, the correlation captured improves quickly in the first a few iterations and tends to converge over iterations.

Feature correlations: Next, we examine whether the studied methods produce same-set and cross-set uncorrelated features. Figures 3(b) and 3(c) plot the average correlations of the first 20 features with all the other features in the same set and in the other set, respectively, *excluding* correlations between the same pairs. *Correlations are nonzero* in both cases for 2D-CCA. In contrast, CCA features are *uncorrelated* in both cases, while MCCA1 and MCCA2 have achieved *same-set and cross-set zero-correlations*, respectively, as supported by derivation in Sec. 3.2. Furthermore, the same-set correlations for MCCA2 and the cross-set correlations for MCCA1 are both low, especially for the first few features. This indicates that they may have similar performance.

Learning and matching: Finally, we investigate how well

[3]Code by D. Hardoon: http://www.davidroihardoon.com/Professional/Code_files/cca.m

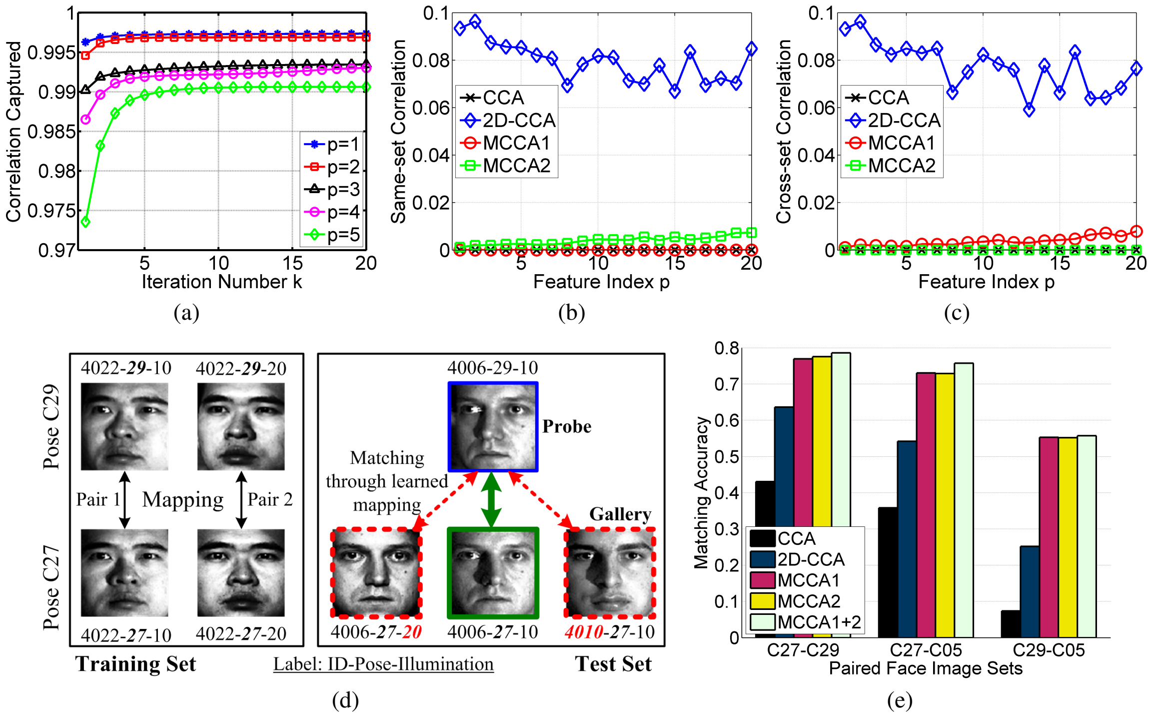

Figure 3: Evaluation on mapping faces of different poses. (a) Convergence study: captured correlations over iterations for the first five MCCA1 features (indexed by p). (b) Validation of same-set zero-correlation for MCCA1 features. (c) Validation of cross-set zero-correlation for MCCA2 features. (d) Experimental design: the numeric codes associated with each face image indicate ID-Pose-Illumination. Correct match is enclosed by solid box and incorrect matches are enclosed by dashed boxes. (e) Matching accuracy comparison, where MCCA1+2 is the fusion of MCCA1 and MCCA2.

MCCA can learn a mapping between face images of different poses. We divide the PIE subset into training set and test set and perform 10-fold cross validation with respect to subjects. Thus, the training set contains Pose1-Pose2 pairs of 61/62 subjects and the test set contains Pose1-Pose2 pairs of the rest 7/6 subjects (no overlapping). We study three pose pairings: C27-C29, C27-C05, and C29-C05. Figure 3(d) shows the design of this experiment.

For each Pose1-Pose2 pairing, we evaluate the matching performance on the test set after training. The gallery consists of $G(= 108/126)$ images of Pose1, projected to $\{\mathbf{r}_g\}$, $g = 1, ..., G$, through learned Pose1 subspace. The probe consists of the same number of Pose2 images. Each probe image is projected to a vector $\mathbf{s}$ in learned Pose2 subspace. With learned canonical correlations $\{\rho_p\}$, we predict the mapping of $\mathbf{s}$ to Pose1 as $\hat{\mathbf{r}} \in \mathbb{R}^P$, where its pth element is $\hat{r}_p = \rho_p s_p$ [Weenink, 2003]. The best match is then $\hat{g} = \arg\min_g dist(\hat{\mathbf{r}}, \mathbf{r}_g)$, where $dist(\cdot)$ is a distance measure. We test three measures to find the best results: L_1 distance, L_2 distance, and angle between feature vectors [Lu *et al.*, 2008]. A *correct* match is found when the probe and gallery images *differ only in pose*. The matching is *incorrect* even if the probe and gallery images are of the same subject but of different illuminations, as illustrated in Fig. 3(d). We report the best results from testing *up to 32 features* for MCCA and *up to 1024 features* for CCA and 2D-CCA.

Figure 3(e) depicts the matching results on three pose pairings. Despite having ***much fewer features*** (only $\frac{32}{1024} = \frac{1}{32}$), MCCA-based methods give the best matching performance for all three scenarios, showing the effectiveness of uncorrelated features and direct learning of compact representations from tensors. There is a small improvement by fusing MCCA1 and MCCA2, implying that they are slightly complementary.

5 Conclusions

We proposed a multilinear CCA method to extract uncorrelated features directly from tensors via tensor-to-vector projection for paired tensor sets. The algorithm successively maximizes correlation captured by each projection pair while enforcing the zero-correlation constraint under two architectures. One architecture produces same-set uncorrelated features while the other produces cross-set uncorrelated features (except the paired features). The solution is iterative and alternates between modes. We evaluated MCCA on learning the mapping between faces of different poses. The results show that it outperforms CCA and 2D-CCA using much fewer features under both architectures. In addition, the two architectures complement each other slightly.

Acknowledgments

The author would like to thank the anonymous reviewers of this work for their very helpful comments.

References

[Anderson, 2003] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, third edition, 2003.

[Chaudhuri *et al.*, 2009] K. Chaudhuri, S. M. Kakade, K. Livescu, and K. Sridharan. Multi-view clustering via canonical correlation analysis. In *Proc. Int. Conf. on Machine Learning*, pages 129–136, 2009.

[Dhillon *et al.*, 2011] P. S. Dhillon, D. Foster, and L. Ungar. Multi-view learning of word embeddings via CCA. In *Advances in Neural Information Processing Systems (NIPS)*, pages 199–207, 2011.

[Gang *et al.*, 2011] L. Gang, Z. Yong, Y-L. Liu, and D. Jing. Three dimensional canonical correlation analysis and its application to facial expression recognition. In *Proc. Int. Conf. on intelligent computing and information science*, pages 56–61, 2011.

[Hardoon *et al.*, 2004] D. Hardoon, S. Szedmak, and J. Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural Computation*, 16(12):2639–2664, 2004.

[Harshman, 1970] R. A. Harshman. Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 16:1–84, 1970.

[Harshman, 2006] R. Harshman. Generalization of canonical correlation to N-way arrays. In *Poster at the 34th Ann. Meeting of the Statistical Soc. Canada*, May 2006.

[He *et al.*, 2005] X. He, D. Cai, and P. Niyogi. Tensor subspace analysis. In *NIPS 18*, pages 499–506, 2005.

[Hotelling, 1936] H. Hotelling. Relations between two sets of variables. *Biometrika*, 28(7):312–377, 1936.

[Kim and Cipolla, 2009] T.-K. Kim and R. Cipolla. Canonical correlation analysis of video volume tensors for action categorization and detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(8):1415–1428, 2009.

[Kolda and Bader, 2009] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, September 2009.

[Kolda and Sun, 2008] T. G. Kolda and J. Sun. Scalable tensor decompositions for multi-aspect data mining. In *Proc. IEEE Int. Conf. on Data Mining*, pages 363–372, December 2008.

[Lee and Choi, 2007] S. H. Lee and S. Choi. Two-dimensional canonical correlation analysis. *IEEE Signal Processing Letters*, 14(10):735738, October 2007.

[Lu *et al.*, 2008] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. MPCA: Multilinear principal component analysis of tensor objects. *IEEE Transactions on Neural Networks*, 19(1):18–39, 2008.

[Lu *et al.*, 2009a] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition. *IEEE Transactions on Neural Networks*, 20(1):103–123, 2009.

[Lu *et al.*, 2009b] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. Uncorrelated multilinear principal component analysis for unsupervised multilinear subspace learning. *IEEE Transactions on Neural Networks*, 20(11):1820–1836, 2009.

[Lu *et al.*, 2011] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. A survey of multilinear subspace learning for tensor data. *Pattern Recognition*, 44(7):1540–1551, July 2011.

[Rai and Daumé III, 2009] P. Rai and H. Daumé III. Multi-label prediction via sparse infinite CCA. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1518–1526, 2009.

[Sim *et al.*, 2003] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, December 2003.

[Sun *et al.*, 2008] L. Sun, S. Ji, and J. Ye. A least squares formulation for canonical correlation analysis. In *Proc. Int. Conf. on Machine Learning*, pages 1024–1031, July 2008.

[Tao *et al.*, 2007] D. Tao, X. Li, X. Wu, and S. J. Maybank. General tensor discriminant analysis and gabor features for gait recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(10):1700–1715, October 2007.

[Wang, 2010] H. Wang. Local two-dimensional canonical correlation analysis. *IEEE Signal Processing Letters*, 17(11):921–924, November 2010.

[Weenink, 2003] D. J. M. Weenink. Canonical correlation analysis. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, 25:8199, 2003.

[Yan *et al.*, 2005] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H.-J Zhang. Discriminant analysis with tensor representation. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 526–532, June 2005.

[Yan *et al.*, 2012] J. Yan, W. Zheng, X. Zhou, and Z. Zhao. Sparse 2-D canonical correlation analysis via low rank matrix approximation for feature extraction. *IEEE Signal Processing Letters*, 19(1):5154, January 2012.

[Ye *et al.*, 2004a] J. Ye, R. Janardan, and Q. Li. GPCA: An efficient dimension reduction scheme for image compression and retrieval. In *ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pages 354–363, 2004.

[Ye *et al.*, 2004b] J. Ye, R. Janardan, and Q. Li. Two-dimensional linear discriminant analysis. In *NIPS 17*, pages 1569–1576, 2004.

[Zhang *et al.*, 2011] Y. Zhang, G. Zhou, Q. Zhao, A. Onishi, J. Jin, X. Wang, and A. Cichocki. Multiway canonical correlation analysis for frequency components recognition in SSVEP-based BCIs. In *Proc. Int. Conf. on Neural Information Processing (ICONIP)*, pages 287–295, 2011.