Robust Multilinear Principal Component Analysis

Kohei Inoue, Kenji Hara, and Kiichi Urahama
Department of Visual Communication Design, Kyushu University
4-9-1, Shiobaru, Minami-ku, Fukuoka, 815-8540, Japan
{k-inoue,hara,urahama}@design.kyushu-u.ac.jp

Abstract

We propose two methods for robustifying multilinear principal component analysis (MPCA) which is an extension of the conventional PCA for reducing the dimensions of vectors to higher-order tensors. For two kinds of outliers, i.e., sample outliers and intra-sample outliers, we derive iterative algorithms on the basis of the Lagrange multipliers. We also demonstrate that the proposed methods outperform the original MPCA when datasets contain such outliers experimentally.

1. Introduction

Dimensionality reduction is of fundamental importance in image processing, pattern recognition and computer vision (CV) as it is necessary to avoid the curse of dimensionality [8]. Principal component analysis (PCA) [12] is one of the most fundamental and well-known techniques for dimensionality reduction. There is a broad community in CV and image analysis that makes use of PCA and its extensions, which includes many application areas, e.g., shape analysis [18], shape statistics [4], segmentation using deformable models [17] and handwritten digit classification [16].

Xu and Yuille [20] proposed self-organizing rules for robust PCA. Higuchi and Eguchi [11] generalized their approach [20] and adaptively selected a set of tuning parameters which control the degree of robustness. De la Torre and Black [5, 6] pointed out that the previous robust approaches to PCA are of limited application in CV problems as they reject entire images as outliers, i.e., they are robust to *sample outliers*, and proposed another robust PCA which is robust to *intra-sample outliers*, e.g., outlier pixels in an image sample. Although these methods achieve improved tolerance to outliers compared with the conventional PCA, it might be difficult to apply them to higher-order tensors because they are formulated for reducing the dimensions of vectors. Therefore, tensors have to be transformed into vectors in advance. However, this transformation might cause a high-dimensional situation in which the above robust PCA approaches fail.

In order to address the problem of high-dimensionality caused by the transformation of tensors into vectors, Yang et al. [21] proposed two-dimensional PCA (2DPCA) for reducing the dimensions of matrices rather than vectors, i.e., matrices do not need to be preliminarily transformed into vectors. However, the 2DPCA is approximately equivalent to the conventional PCA operated only on the row vectors of matrices [19, 9, 25]. Ye et al. [23] proposed generalized PCA (GPCA) which reduces the dimensions of both rows and columns of matrices, i.e., GPCA is a generalization of 2DPCA [21]. Cai et al. [2] proposed Tensor PCA and Tensor LDA; the former is essentially equivalent to GPCA [23]. Ye [22] also proposed generalized low rank approximations of matrices (GLRAM) which is highly similar to GPCA [23]. While GPCA [23] assumes that data are of zero mean, GLRAM [22] does not. Recently, Lu et al. [13] proposed multilinear PCA (MPCA) which reduces the dimensions of all modes of higher-order tensors, i.e., MPCA is a generalization of GPCA [23] and 2DPCA [21]. Although these extensions of PCA for vectors to that for higher-order tensors resolve the problem resulting from the high-dimensionality of vectors, their tolerance to outliers is limited because they are formulated as Frobenius norm minimization problems.

This work is motivated by the above observation of the state-of-the-art PCA techniques. That is, robustifying MPCA [13] will improve the tolerance of MPCA to outliers and avoid the problem of high-dimensionality simultaneously. Since the conventional PCA and its extensions including the above techniques are widely used in CV and other areas, the robust version of MPCA will increase the applicability of the PCA techniques in various areas.

In this paper, we propose two methods for robustifying the MPCA [13] to sample outliers and intra-sample outliers. For both of them, we derive iterative algorithms on the basis of the Lagrange multipliers. We experimentally compare our robust MPCA (RMPCA) with MPCA [13] and show the effectiveness of the RMPCA.

The rest of this paper is organized as follows. Section

591

2 summarizes the MPCA [13]. Section 3 describes our RMPCA for both sample outliers and intra-sample outliers. Section 4 illustrates experimental results. Finally, Section 5 concludes this paper.

2. Multilinear Principal Component Analysis

In this section, we summarize the multilinear principal component analysis (MPCA) [13] and rewrite the objective function for the MPCA to another form which is suitable for the robust one described in the subsequent section.

Let $\{\mathcal{A}_1, \ldots, \mathcal{A}_M\}$ be a set of tensors, where $\mathcal{A}_m = [a_{m i_1 \ldots i_N}] \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ for $m = 1, \ldots, M$ is an Nth-order tensor and $a_{m i_1 \ldots i_N}$ is the $(i_1, \ldots, i_N)$ element of $\mathcal{A}_m$. The total scatter of these tensors is defined as $\Psi_{\mathcal{A}} = \sum_{m=1}^{M} \| \mathcal{A}_m - \bar{\mathcal{A}} \|_F^2$, where $\bar{\mathcal{A}} = [\bar{a}_{i_1 \ldots i_N}] = \frac{1}{M} \sum_{m=1}^{M} \mathcal{A}_m$ and $\| \cdot \|_F$ denotes the Frobenius norm. Let $U = \{U^{(1)}, \ldots, U^{(N)}\}$ be a set of matrices, where $U^{(n)} = [u_{i_n j_n}^{(n)}] \in \mathbb{R}^{I_n \times J_n}$, $J_n \leq I_n$ for $n = 1, \ldots, N$ and $u_{i_n j_n}^{(n)}$ is the (i_n, j_n) element of $U^{(n)}$. Then the sequence of the n-mode product [7, 1] of $\mathcal{A}_m$ and $U^T = \{U^{(1)T}, \ldots, U^{(N)T}\}$, where $U^{(n)T}$ denotes the transpose of $U^{(n)}$, is denoted by

$$
\begin{aligned}
\mathcal{B}_m &= \mathcal{A}_m \times \{U^T\} & (1) \\
&= \mathcal{A}_m \times_1 U^{(1)T} \cdots \times_N U^{(N)T}, & (2)
\end{aligned}
$$

where $\mathcal{B}_m = [b_{m j_1 \ldots j_N}] \in \mathbb{R}^{J_1 \times \cdots \times J_N}$ and $b_{m j_1 \ldots j_N}$ is the $(j_1, \ldots, j_N)$ element of $\mathcal{B}_m$, and $\mathcal{A}_m \times_n U^{(n)T} = [\sum_{i_n=1}^{I_n} a_{m i_1 \ldots i_n \ldots i_N} u_{i_n j_n}^{(n)}] \in \mathbb{R}^{I_1 \times \cdots I_{n-1} \times J_n \times I_{n+1} \times \cdots \times I_N}$ is the n-mode product of $\mathcal{A}_m$ and $U^{(n)T}$. The MPCA is formulated as follows:

$$
\begin{aligned}
& \max_U \quad \Psi_{\mathcal{B}} & (3) \\
& \text{subj.to} \quad U^{(n)T} U^{(n)} = I_{J_n}, \quad n = 1, \ldots, N, & (4)
\end{aligned}
$$

where $\Psi_{\mathcal{B}} = \sum_{n=1}^{M} \| \mathcal{B}_m - \bar{\mathcal{B}} \|_F^2$, $\bar{\mathcal{B}} = \frac{1}{M} \sum_{m=1}^{M} \mathcal{B}_m$ is the total scatter of $\{\mathcal{B}_1, \ldots, \mathcal{B}_M\}$ and I_{J_n} is the $J_n \times J_n$ identity matrix.

Since $\Psi_{\mathcal{A}}$ is a constant, we may rewrite (3) as

$$
\min_U \quad \Psi_{\mathcal{A}} - \Psi_{\mathcal{B}}, \tag{5}
$$

from which we have another objective function for the MPCA as follows:

$$
\begin{aligned}
& \Psi_{\mathcal{A}} - \Psi_{\mathcal{B}} & (6) \\
&= \sum_{m=1}^{M} \left\| \tilde{\mathcal{A}}_m \right\|_F^2 - \sum_{m=1}^{M} \left\| \tilde{\mathcal{B}}_m \right\|_F^2 & (7) \\
&= \sum_{m=1}^{M} \left\| \tilde{\mathcal{A}}_m \right\|_F^2 - \sum_{m=1}^{M} \left\| \tilde{\mathcal{A}}_m \times \{U^T\} \right\|_F^2 & (8) \\
&= \sum_{m=1}^{M} \left(\left\| \tilde{\mathcal{A}}_m \right\|_F^2 - 2 \left\| \tilde{\mathcal{A}}_m \times \{U^T\} \right\|_F^2 \right. \\
& \qquad \left. + \left\| \tilde{\mathcal{A}}_m \times \{U^T\} \right\|_F^2 \right) & (9) \\
&= \sum_{m=1}^{M} \left\| \tilde{\mathcal{A}}_m - \left(\tilde{\mathcal{A}}_m \times \{U^T\} \right) \times \{U\} \right\|_F^2 & (10) \\
&= \sum_{m=1}^{M} \left\| \tilde{\mathcal{A}}_m - \tilde{\mathcal{B}}_m \times \{U\} \right\|_F^2 & (11) \\
&= \sum_{m=1}^{M} \left\| \mathcal{A}_m - \bar{\mathcal{A}} - \tilde{\mathcal{B}}_m \times \{U\} \right\|_F^2, & (12)
\end{aligned}
$$

where $\tilde{\mathcal{A}}_m = [\tilde{a}_{m i_1 \ldots i_N}] = \mathcal{A}_m - \bar{\mathcal{A}}$ and $\tilde{\mathcal{B}}_m = [\tilde{b}_{m j_1 \ldots j_N}] = \mathcal{B}_m - \bar{\mathcal{B}}$.

Unlike in the above Frobenius norm minimization formulation where the mean tensor $\bar{\mathcal{A}}$ is obtained analytically, in the following robust MPCA (RMPCA), $\bar{\mathcal{A}}$, i.e., the *robust mean*, must be estimated as well as $\{\tilde{\mathcal{B}}\}$ and U.

3. Robust MPCA

The above MPCA is formulated as a minimization of the sum of the Frobenius norms. Therefore, the MPCA is not robust to outliers. In this section, we propose two methods for robustifying the MPCA to sample outliers and intra-sample outliers.

3.1. RMPCA for Sample Outliers

First, we modify the minimization of (12) as

$$
\min_{\bar{\mathcal{A}}, \{\tilde{\mathcal{B}}_m\}_{m=1}^{M}, U} \sum_{m=1}^{M} \rho \left(\left\| \mathcal{A}_m - \bar{\mathcal{A}} - \tilde{\mathcal{B}}_m \times \{U\} \right\|_F \right), \tag{13}
$$

where $\rho(x)$ is a ρ-function for robust M-estimation. Among several M-estimators including Andrews, bisquare, Cauchy, and Huber estimators [10], we select the Welsch estimator for which the ρ-function is given by

$$
\rho(x) = 1 - e^{-\alpha x^2}, \tag{14}
$$

because it worked well in our experiments described later. Let $F(\bar{\mathcal{A}}, \{\tilde{\mathcal{B}}_m\}, U)$ be the objective function in (13) with

(14). Then we have

$$F\left(\bar{\mathcal{A}}, \{\tilde{\mathcal{B}}_m\}, U\right) = M - \sum_{m=1}^{M} e^{-\alpha\left\|\mathcal{A}_m - \bar{\mathcal{A}} - \tilde{\mathcal{B}}_m \times \{U\}\right\|_F^2}. \tag{15}$$

Since M is a constant, we may rewrite (13) as

$$\max_{\bar{\mathcal{A}}, \{\tilde{\mathcal{B}}_m\}_{m=1}^{M}, U} \tilde{F}\left(\bar{\mathcal{A}}, \{\tilde{\mathcal{B}}_m\}, U\right), \tag{16}$$

where

$$\tilde{F}\left(\bar{\mathcal{A}}, \{\tilde{\mathcal{B}}_m\}, U\right) = M - F\left(\bar{\mathcal{A}}, \{\tilde{\mathcal{B}}_m\}, U\right) \tag{17}$$

$$= \sum_{m=1}^{M} e^{-\alpha\left\|\mathcal{A}_m - \bar{\mathcal{A}} - \tilde{\mathcal{B}}_m \times \{U\}\right\|_F^2}. \tag{18}$$

The Lagrange function for (16) with (4) is given by

$$L\left(\bar{\mathcal{A}}, \{\tilde{\mathcal{B}}_m\}, U, \{\Lambda^{(n)}\}\right) = \tilde{F}\left(\bar{\mathcal{A}}, \{\tilde{\mathcal{B}}_m\}, U\right)$$
$$+ \alpha \sum_{n=1}^{N} \mathrm{tr}\left[\Lambda^{(n)}\left(U^{(n)^T} U^{(n)} - I_{J_n}\right)\right], \tag{19}$$

where tr denotes the matrix trace and $\Lambda^{(n)} \in \mathbb{R}^{J_n \times J_n}$ for $n = 1, \ldots, N$ is a symmetric matrix of which the elements are the Lagrange multipliers. Then we have the following necessary conditions for optimality:

$$\frac{\partial L}{\partial \bar{\mathcal{A}}} = 2\alpha \sum_{m=1}^{M} \left(\mathcal{C}_m - \bar{\mathcal{A}}\right) e^{-\alpha\left\|\mathcal{C}_m - \bar{\mathcal{A}}\right\|_F^2} = 0, \tag{20}$$

where $\mathcal{C}_m = \mathcal{A}_m - \tilde{\mathcal{B}}_m \times \{U\}$, and

$$\frac{\partial L}{\partial \tilde{\mathcal{B}}_m} = 2\alpha \left(\tilde{\mathcal{A}}_m - \tilde{\mathcal{B}}_m \times \{U\}\right) \times \{U^T\}$$
$$e^{-\alpha\left\|\tilde{\mathcal{A}}_m - \tilde{\mathcal{B}}_m \times \{U\}\right\|_F^2} = 0 \tag{21}$$

for $m = 1, \ldots, M$, and

$$\frac{1}{2\alpha} \frac{\partial L}{\partial U^{(n)}} = \sum_{m=1}^{M} \left[\tilde{A}_{m(n)} - U^{(n)} \tilde{B}_{m(n)}^{(-n)}\right] \left(\tilde{B}_{m(n)}^{(-n)}\right)^T$$
$$e^{-\alpha\left\|\tilde{\mathcal{A}}_m - \tilde{\mathcal{B}}_m \times \{U\}\right\|_F^2} + U^{(n)} \Lambda^{(n)} = 0 \tag{22}$$

for $n = 1, \ldots, N$, where $\tilde{A}_{m(n)}$ and $\tilde{B}_{m(n)}^{(-n)}$ are the mode-n matricizing [1] of $\tilde{\mathcal{A}}_m$ and $\tilde{\mathcal{B}}_m^{(-n)} = \tilde{\mathcal{B}}_m \times_{-n} \{U\} = \tilde{\mathcal{B}}_m \times_1 U^{(1)} \cdots \times_{n-1} U^{(n-1)} \times_{n+1} U^{(n+1)} \cdots \times_N U^{(N)}$, respectively. We also have

$$\frac{1}{\alpha} \frac{\partial L}{\partial \Lambda^{(n)}} = U^{(n)^T} U^{(n)} - I_{J_n} = 0 \tag{23}$$

for $n = 1, \ldots, N$.

From (20) and (21), we have

$$\bar{\mathcal{A}} = \frac{\sum_{m=1}^{M} \mathcal{C}_m e^{-\alpha\left\|\mathcal{C}_m - \bar{\mathcal{A}}\right\|_F^2}}{\sum_{m=1}^{M} e^{-\alpha\left\|\mathcal{C}_m - \bar{\mathcal{A}}\right\|_F^2}} \tag{24}$$

and

$$\tilde{\mathcal{B}}_m = \tilde{\mathcal{A}}_m \times \{U^T\}, \tag{25}$$

respectively. From (22), we have

$$U^{(n)} = P^{(n)} \left(Q^{(n)} - \Lambda^{(n)}\right)^{-1}, \tag{26}$$

where

$$P^{(n)} = \sum_{m=1}^{M} \tilde{A}_{m(n)} \left(\tilde{B}_{m(n)}^{(-n)}\right)^T e^{-\alpha\left\|\tilde{\mathcal{A}}_m - \tilde{\mathcal{B}}_m \times \{U\}\right\|_F^2}, \tag{27}$$

$$Q^{(n)} = \sum_{m=1}^{M} \tilde{B}_{m(n)}^{(-n)} \left(\tilde{B}_{m(n)}^{(-n)}\right)^T e^{-\alpha\left\|\tilde{\mathcal{A}}_m - \tilde{\mathcal{B}}_m \times \{U\}\right\|_F^2}. \tag{28}$$

By substituting (26) into (23) we find that

$$P^{(n)^T} P^{(n)} = \left(Q^{(n)} - \Lambda^{(n)}\right)^T \left(Q^{(n)} - \Lambda^{(n)}\right). \tag{29}$$

Let

$$P^{(n)^T} P^{(n)} = V^{(n)} \Sigma^{(n)} V^{(n)^T} \tag{30}$$

be a spectral decomposition of $P^{(n)^T} P^{(n)}$, where $\Sigma^{(n)}$ is a diagonal matrix of which the diagonal elements are the eigenvalues of $P^{(n)^T} P^{(n)}$ and $V^{(n)}$ is an orthogonal matrix of which the columns are the corresponding eigenvectors of $P^{(n)^T} P^{(n)}$. Then it follows from (29) and (30) that

$$\Lambda^{(n)} = Q^{(n)} - \Sigma^{(n)1/2} V^{(n)^T}. \tag{31}$$

Substituting this into (26), we have

$$U^{(n)} = P^{(n)} V^{(n)} \Sigma^{(n)-1/2}. \tag{32}$$

Consequently, we obtain the following algorithm:
Algorithm 1.

Step 0: Initialize $\bar{\mathcal{A}}$ and U as $\bar{\mathcal{A}}^{(0)} = \frac{1}{M} \sum_{m=1}^{M} \mathcal{A}_m$ and $U^{(0)} = \{U^{(1,0)}, \ldots, U^{(N,0)}\}$, where $U^{(n,0)}$ is an orthogonal matrix of which the columns are the eigenvectors of $\sum_{m=1}^{M} \tilde{A}_{m(n)}^{(0)} (\tilde{A}_{m(n)}^{(0)})^T$ corresponding to the largest J_n eigenvalues, where $\tilde{A}_{m(n)}^{(0)}$ is the mode-n matricizing of $\tilde{\mathcal{A}}_m^{(0)} = \mathcal{A}_m - \bar{\mathcal{A}}^{(0)}$. Then initialize $\tilde{\mathcal{B}}_m$ as $\tilde{\mathcal{B}}_m^{(0)} = \tilde{\mathcal{A}}_m^{(0)} \times \{U^{(0)^T}\}$ for $m = 1, \ldots, M$. Initialize the number of iterations as $t = 0$.

Step 1: Compute $U^{(t+1)} = \{U^{(1,t+1)}, \ldots, U^{(N,t+1)}\}$, where $U^{(n,t+1)}$ is given by (32) into which $\bar{\mathcal{A}}^{(t)}$, $U^{(t)}$ and $\{\tilde{\mathcal{B}}_m^{(t)}\}$ are substituted.

Step 2: For $m = 1, \ldots, M$, compute $\tilde{\mathcal{B}}_m^{(t+1)}$ by (25) into which $\bar{\mathcal{A}}^{(t)}$ and $U^{(t+1)}$ are substituted.

Step 3: Compute $\bar{A}^{(t+1)}$ by (24) into which $\bar{\mathcal{A}}^{(t)}$, $\{\tilde{\mathcal{B}}_m^{(t+1)}\}$ and $U^{(t+1)}$ are substituted.

Step 4: If, for $\epsilon > 0$, $[\tilde{F}(\bar{A}^{(t+1)}, \{\tilde{\mathcal{B}}_m^{(t+1)}\}, U^{(t+1)}) - \tilde{F}(\bar{A}^{(t)}, \{\tilde{\mathcal{B}}_m^{(t)}\}, U^{(t)})]/M < \epsilon$, then stop, otherwise increase t by 1 and go to Step 1.

3.2. RMPCA for Intra-Sample Outliers

Next, we modify the minimization of (12) as

$$\min_{\bar{A}, \{\tilde{\mathcal{B}}_m\}_{m=1}^M, U} \sum_{m=1}^{M} \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} \rho\left(d_{mi_1 \ldots i_N}\right), \quad (33)$$

where $d_{mi_1 \ldots i_N} = a_{mi_1 \ldots i_N} - \bar{a}_{i_1 \ldots i_N} - \sum_{j_1=1}^{J_1} \cdots \sum_{j_N=1}^{J_N} \tilde{b}_{mj_1 \ldots j_N} u_{i_1 j_1}^{(1)} \cdots u_{i_N j_N}^{(N)}$. Let $G(\bar{A}, \{\tilde{\mathcal{B}}_m\}, U)$ be the objective function in (33) with (14). Then we have

$$G\left(\bar{A}, \{\tilde{\mathcal{B}}_m\}, U\right) = \Theta - \sum_{m=1}^{M} \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} e^{-\alpha d_{mi_1 \ldots i_N}^2}, \quad (34)$$

where $\Theta = M \prod_{n=1}^{N} I_n$. Since Θ is a constant, we may rewrite (33) as

$$\max_{\bar{A}, \{\tilde{\mathcal{B}}_m\}_{m=1}^M, U} \tilde{G}\left(\bar{A}, \{\tilde{\mathcal{B}}_m\}, U\right), \quad (35)$$

where

$$\tilde{G}\left(\bar{A}, \{\tilde{\mathcal{B}}_m\}, U\right) = \Theta - G\left(\bar{A}, \{\tilde{\mathcal{B}}_m\}, U\right) \quad (36)$$

$$= \sum_{m=1}^{M} \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} e^{-\alpha d_{mi_1 \ldots i_N}^2}. \quad (37)$$

The Lagrange function for (35) with (4) is given by

$$\tilde{L}\left(\bar{A}, \{\tilde{\mathcal{B}}_m\}, U, \{\tilde{\Lambda}^{(n)}\}\right) = \tilde{G}\left(\bar{A}, \{\tilde{\mathcal{B}}_m\}, U\right)$$
$$+ \alpha \sum_{n=1}^{N} \operatorname{tr}\left[\tilde{\Lambda}^{(n)}\left(U^{(n)T} U^{(n)} - I_{J_n}\right)\right], \quad (38)$$

where $\tilde{\Lambda}^{(n)} \in \mathbb{R}^{J_n \times J_n}$ for $n = 1, \ldots, N$ is a symmetric matrix of which the elements are the Lagrange multipliers. Then we have the following necessary conditions for optimality:

$$\frac{\partial \tilde{L}}{\partial \bar{\mathcal{A}}} = 2\alpha \sum_{m=1}^{M} \left(\mathcal{C}_m - \bar{\mathcal{A}}\right) \odot \mathcal{E}_m = 0, \quad (39)$$

where $\mathcal{E}_m = [e^{-\alpha d_{mi_1 \ldots i_N}^2}] \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ for $m = 1, \ldots, M$ and $\odot$ denotes the Hadamard product or elementwise product [14] of tensors, and

$$\frac{\partial \tilde{L}}{\partial \tilde{\mathcal{B}}_m} = 2\alpha \left[\left(\tilde{\mathcal{A}}_m - \tilde{\mathcal{B}}_m \times \{U\}\right) \odot \mathcal{E}_m\right] \times \{U^T\} = 0 \quad (40)$$

for $m = 1, \ldots, M$, and

$$\frac{1}{2\alpha} \frac{\partial \tilde{L}}{\partial U^{(n)}} = \sum_{m=1}^{M} \left\langle \left(\tilde{\mathcal{A}}_m - \tilde{\mathcal{B}}_m \times \{U\}\right) \odot \mathcal{E}_m, \right.$$
$$\left. \tilde{\mathcal{B}}_m^{(-n)} \right\rangle_{\{K;K\}} + U^{(n)} \tilde{\Lambda}^{(n)} = 0 \quad (41)$$

for $n = 1, \ldots, N$, where $K = \{1, \ldots, n-1, n+1, \ldots, N\}$ and $\langle \cdot, \cdot \rangle_{\{\ldots;\ldots\}}$ denotes the contracted product [1] of tensors, and

$$\frac{1}{\alpha} \frac{\partial \tilde{L}}{\partial \tilde{\Lambda}^{(n)}} = U^{(n)T} U^{(n)} - I_{J_n} = 0 \quad (42)$$

for $n = 1, \ldots, N$.

From (39) and (40), we have

$$\bar{A} = \left(\sum_{m=1}^{M} \mathcal{C}_m \odot \mathcal{E}_m\right) \odot \mathcal{H}, \quad (43)$$

where $\mathcal{H} = [1/\sum_{m=1}^{M} e^{-\alpha d_{mi_1 \ldots i_N}^2}] \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, and (25), respectively. From (41), we have

$$U^{(n)} = \tilde{P}^{(n)} \left(\tilde{Q}^{(n)} - \tilde{\Lambda}^{(n)}\right)^{-1}, \quad (44)$$

where

$$\tilde{P}^{(n)} = \sum_{m=1}^{M} \left\langle \tilde{\mathcal{A}}_m \odot \mathcal{E}_m, \tilde{\mathcal{B}}_m^{(-n)} \right\rangle_{\{K;K\}}, \quad (45)$$

$$\tilde{Q}^{(n)} = \sum_{m=1}^{M} \left\langle \tilde{\mathcal{B}}_m^{(-n)} \odot \tilde{\mathcal{B}}_m^{(-n)}, \mathcal{E}_m \right\rangle_{\{K;K\}}. \quad (46)$$

Since (44) is analogous to (26), we can derive the following equation in the same manner as the derivation of (32):

$$U^{(n)} = \tilde{P}^{(n)} \tilde{V}^{(n)} \left(\tilde{\Sigma}^{(n)}\right)^{-1/2}, \quad (47)$$

where $\tilde{\Sigma}^{(n)}$ is a diagonal matrix of which the diagonal elements are the eigenvalues of $(\tilde{P}^{(n)})^T \tilde{P}^{(n)}$ and $\tilde{V}^{(n)}$ is an orthogonal matrix of which the columns are the corresponding eigenvectors of $(\tilde{P}^{(n)})^T \tilde{P}^{(n)}$.

Consequently, we obtain the following algorithm:

Algorithm 2.

Figure 1. Input images: the first five images of the 27th person in the ORL database [15] and a sample outlier.

Step 0: Initialize $\bar{\mathcal{A}}$ and U as $\bar{\mathcal{A}}^{(0)} = \frac{1}{M}\sum_{m=1}^{M}\mathcal{A}_m$ and $U^{(0)} = \{U^{(1,0)}, \dots, U^{(N,0)}\}$, where $U^{(n,0)}$ is an orthogonal matrix of which the columns are the eigenvectors of $\sum_{m=1}^{M}\tilde{A}^{(0)}_{m(n)}(\tilde{A}^{(0)}_{m(n)})^T$ corresponding to the largest J_n eigenvalues, where $\tilde{A}^{(0)}_{m(n)}$ is the mode-n matricizing of $\tilde{\mathcal{A}}^{(0)}_m = \mathcal{A}_m - \bar{\mathcal{A}}^{(0)}$. Then initialize $\tilde{\mathcal{B}}_m$ as $\tilde{\mathcal{B}}^{(0)}_m = \tilde{\mathcal{A}}^{(0)}_m \times \{U^{(0)T}\}$ for $m = 1, \dots, M$. Initialize the number of iterations as $t = 0$.

Step 1: Compute $U^{(t+1)} = \{U^{(1,t+1)}, \dots, U^{(N,t+1)}\}$, where $U^{(n,t+1)}$ is given by (47) into which $\bar{\mathcal{A}}^{(t)}$, $U^{(t)}$ and $\{\tilde{\mathcal{B}}^{(t)}_m\}$ are substituted.

Step 2: For $m = 1, \dots, M$, compute $\tilde{\mathcal{B}}^{(t+1)}_m$ by (25) into which $\bar{\mathcal{A}}^{(t)}$ and $U^{(t+1)}$ are substituted.

Step 3: Compute $\bar{\mathcal{A}}^{(t+1)}$ by (43) into which $\bar{\mathcal{A}}^{(t)}$, $\{\tilde{\mathcal{B}}^{(t+1)}_m\}$ and $U^{(t+1)}$ are substituted.

Step 4: If, for $\epsilon > 0$, $[\tilde{G}(\bar{\mathcal{A}}^{(t+1)}, \{\tilde{\mathcal{B}}^{(t+1)}_m\}, U^{(t+1)}) - \tilde{G}(\bar{\mathcal{A}}^{(t)}, \{\tilde{\mathcal{B}}^{(t)}_m\}, U^{(t)})]/\Theta < \epsilon$, then stop, otherwise increase t by 1 and go to Step 1.

4. Experimental Results

In this section, we experimentally evaluate the performance of the proposed algorithms on the ORL face image database [15]. The ORL database [15] contains face images of 40 persons. For each person, there are 10 different face images. That is, the total number of the images in the database is 400. The size of each image is 112×92 pixels, i.e., $I_1 = 112$, $I_2 = 92$ and $N = 2$. An example of face images in the database is shown in Fig. 1, where five images of a person are arranged along with an example of sample outliers (the rightmost image). Reconstructed images with MPCA and RMPCA for sample outliers are shown in Fig. 2. Five reconstructed images from outlier-free data are shown at the top row in Fig. 2(a), i.e., the clean reconstruction, where the rightmost image is the mean image of 10 face images of the person shown in Fig. 1. The corresponding zoomed parts of the forehead of the person are aligned at the bottom row in Fig. 2(a). Reconstructed images with MPCA and RMPCA for data including sample outliers are shown in Figs. 2 (b) and (c), respectively. In Fig. 2(b), the left five reconstructed images and the rightmost mean image
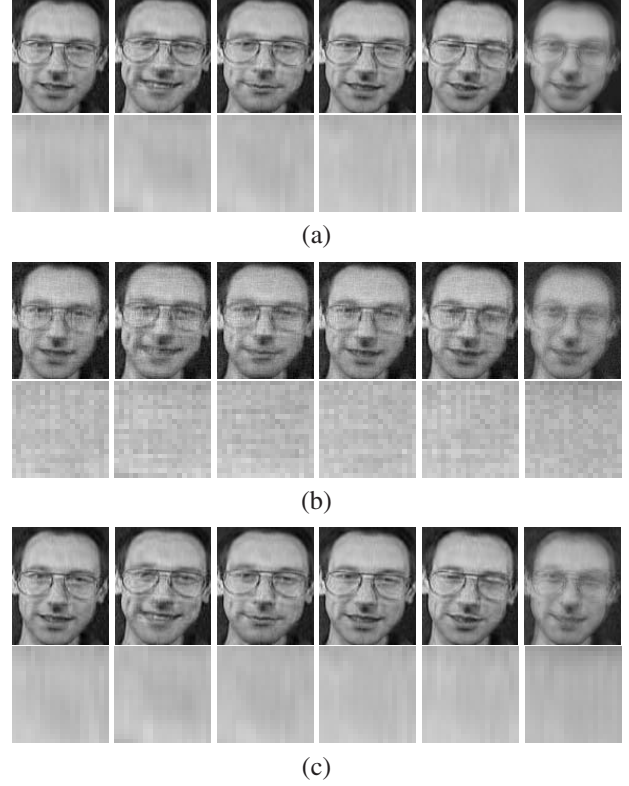

(a)

(b)

(c)

Figure 2. Reconstructed images and (robust) mean images: (a) outlier-free reconstruction by MPCA, (b) MPCA, (c) RMPCA. In each subfigure, whole images are aligned at the top row and the corresponding zoomed parts are aligned at the bottom row.

are disturbed by the sample outlier. On the other hand, in Fig. 2(c), the left five reconstructed images are less sensitive to the outlier and the rightmost robust mean image is close to the outlier-free mean image in Fig. 2(a). In this example, we set $J_1 = J_2 = 30$ for both MPCA and RMPCA, and $\alpha = 10^{-6}$ and $\epsilon = 10^{-6}$ for RMPCA. Although we select the value of α manually, some self-tuning algorithms [3, 24] might work well. Let $\tilde{M}$ be the number of sample outliers per person. Then we set $M = 10 + \tilde{M}$ for each person. Reconstructed images are computed as $\hat{\mathcal{A}}_m = \tilde{\mathcal{B}}_m \times \{U\} + \bar{\mathcal{A}}$ for $m = 1, \dots, 10$. Sample outliers are numbered from 11 to M. Variation in the objective function $\tilde{F}$ until the convergence is shown in Fig. 3, in which the horizontal axis denotes the number of iterations and the vertical axis denotes the value of $\tilde{F}$ at each iteration. Reconstruction errors calculated with all face images in the ORL database [15] are shown in Fig. 4, in which we evaluated the errors by the root mean squared error (RMSE) defined as

$$\text{RMSE} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left\|\mathcal{A}_m - \hat{\mathcal{A}}_m\right\|_F^2}. \qquad (48)$$

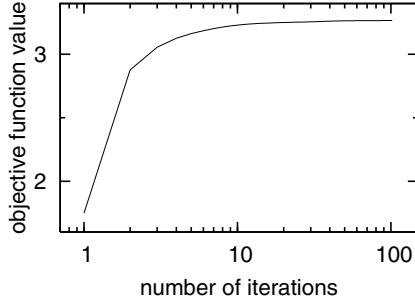In this figure, the horizontal axis denotes the number of
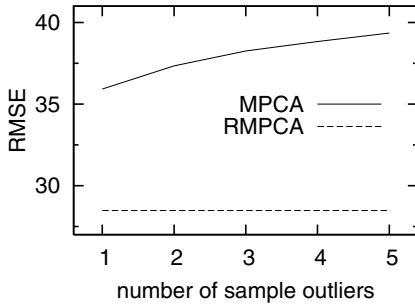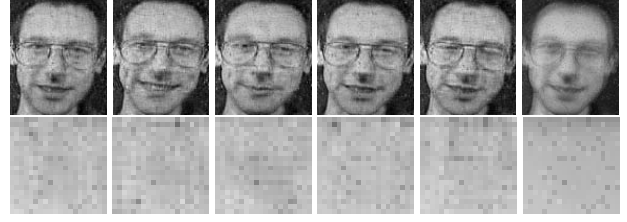
Figure 3. Variation in $\tilde{F}$.
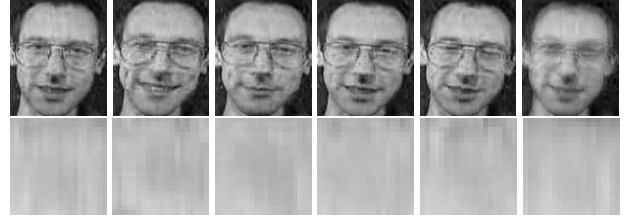

Figure 4. Reconstruction errors.

Figure 5. Input images including intra-sample outliers.

(a)

(b)

Figure 6. Reconstructed images and (robust) mean images: (a) MPCA, (b) RMPCA. In each subfigure, whole images are aligned at the top row and the corresponding zoomed parts are aligned at the bottom row.
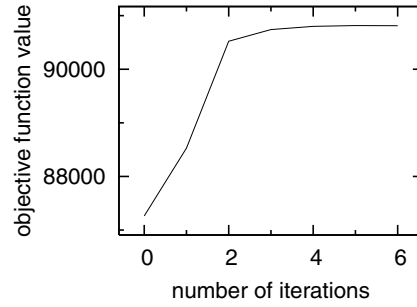

Figure 7. Variation in $\tilde{G}$.

sample outliers per person and the vertical axis denotes the RMSE. MPCA and RMPCA are denoted by solid and broken lines, respectively. The RMSE for RMPCA is lower than that for MPCA. Although the RMSE for MPCA increases with the number of sample outliers, that for RMPCA is almost constant.

An example of face images including intra-sample outliers is shown in Fig. 5. In each image, 2% of pixels are outliers. In this experiment, we use the MATLAB function 'imnoise' with 'salt & pepper' option to generate intra-sample outliers. Reconstructed images with MPCA and RMPCA for intra-sample outliers are shown in Figs. 6 (a) and (b), respectively. At the top row in Fig. 6(a), the left five reconstructed images and the rightmost mean image are disturbed by the intra-sample outliers. The corresponding zoomed parts are aligned at the bottom row. On the other hand, in Fig. 6(b), the left five reconstructed images and the rightmost robust mean image are less sensitive to the outliers. In this example, we set $\alpha = 10^{-3}$ and $M = 10$. Variation in the objective function $\tilde{G}$ until convergence is shown

in Fig. 7, in which the horizontal axis denotes the number of iterations and the vertical axis denotes the value of $\tilde{G}$. Reconstruction errors evaluated by the RMSE in (48) are shown in Fig. 8, in which the horizontal axis denotes the ratio of intra-sample outliers in all pixels and the vertical axis denotes the RMSE. In this figure, MPCA and RMPCA are denoted by solid and broken lines, respectively. The RMSE for RMPCA is lower than that for MPCA. In contrast to the RMSE for RMPCA in Fig. 4, which is almost constant, that in Fig. 8 increases with the ratio of intra-sample outliers. This result demonstrates the difficulty in detecting and removing the intra-sample outliers compared to the sample outliers.

5. Conclusion

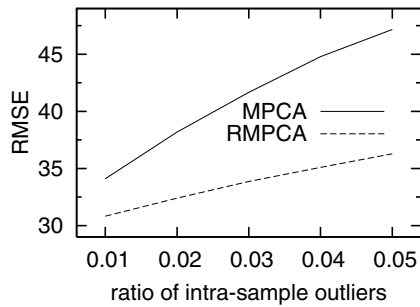We have proposed two robust methods for multilinear principal component analysis (MPCA): one for sample out-

Figure 8. Reconstruction errors.

liers and another for intra-sample outliers. For both of them, we derived iterative algorithms on the basis of the Lagrange multipliers. Experimental results show that the proposed methods outperform the original MPCA when datasets contain outliers. Future work will include the robustification of multilinear discriminant analysis.

Acknowledgments

References

[1] B. W. Bader and T. G. Kolda. Algorithm 862: Matlab tensor classes for fast algorithm prototyping. *ACM Trans. Math. Softw.*, 32(4):635–653, December 2006. 2, 3, 4

[2] D. Cai, X. He, and J. Han. Subspace learning based on tensor analysis. *Department of Computer Science Technical Report No. 2572, University of Illinois at Urbana-Champaign (UIUCDCS-R-2005-2572)*, May 2005. 1

[3] D. Comaniciu. An algorithm for data-driven bandwidth selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):281–288, 2003. 5

[4] D. Cremers, T. Kohlberger, and C. Schnörr. Shape statistics in kernel space for variational image segmentation. *Pattern Recognition*, 36(9):1929–1943, September 2003. 1

[5] F. De la Torre and M. J. Black. Robust principal component analysis for computer vision. In *Proc. Int. Conf. on Computer Vision*, volume 1, pages 362–369, 2001. 1

[6] F. De la Torre and M. J. Black. A framework for robust subspace learning. *International Journal of Computer Vision*, 54(1):117–142, 2003. 1

[7] L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000. 2

[8] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, November 2000. 1

[9] Q.-x. Gao. Is two-dimensional pca equivalent to a special case of modular pca? *Pattern Recogn. Lett.*, 28(10):1250–1251, 2007. 1

[10] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions (Wiley Series in Probability and Statistics)*. Wiley-Interscience, New York, revised edition, April 2005. 2

[11] I. Higuchi and S. Eguchi. Robust principal component analysis with adaptive selection for tuning parameters. *Journal of Machine Learning Research*, 5:453–471, 2004. 1

[12] I. T. Jolliffe. *Principal Component Analysis*. Springer, second edition, October 2002. 1

[13] 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. 1, 2

[14] H. Lütkepohl. *Handbook of Matrices*. John Wiley & Sons Inc, October 1996. 4

[15] F. Samaria and A. Harter. Parameterisation of a stochastic model for human face identification. In *Proc. 2nd IEEE Workshop on Applications of Computer Vision*, 1994. 5

[16] B. Savas and L. Eldén. Handwritten digit classification using higher order singular value decomposition. *Pattern Recogn.*, 40(3):993–1003, 2007. 1

[17] J. Schmid and N. Magnenat-Thalmann. Mri bone segmentation using deformable models and shape priors. In *MICCAI '08: Proceedings of the 11th international conference on Medical Image Computing and Computer-Assisted Intervention - Part I*, pages 119–126, Berlin, Heidelberg, 2008. Springer-Verlag. 1

[18] M. B. Stegmann and D. D. Gomez. A brief introduction to statistical shape analysis, mar 2002. Images, annotations and data reports are placed in the enclosed zip-file. 1

[19] L. Wang, X. Wang, X. Zhang, and J. Feng. The equivalence of two-dimensional pca to line-based pca. *Pattern Recogn. Lett.*, 26(1):57–60, 2005. 1

[20] L. Xu and A. L. Yuille. Robust principal component analysis by self-organizing rules based on statistical physics approach. *IEEE Transactions on Neural Networks*, 6(1):131–143, 1995. 1

[21] J. Yang, D. Zhang, A. Frangi, and J. yu Yang. Two-dimensional pca: a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):131–137, 2004. 1

[22] J. Ye. Generalized low rank approximations of matrices. *Machine Learning*, 61(1-3):167–191, November 2005. 1

[23] J. Ye, R. Janardan, and Q. Li. Gpca: an efficient dimension reduction scheme for image compression and retrieval. In *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 354–363, New York, NY, USA, 2004. ACM Press. 1

[24] L. Zelnik-Manor and P. Perona. Self-tuning spectral clustering. In L. K. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems 17*, pages 1601–1608. MIT Press, Cambridge, MA, 2005. 5

[25] D. Zhang, S. Chen, and J. Liu. Representing image matrices: Eigenimages versus eigenvectors. In *LNCS, Advances in Neural Networks – ISNN 2005*, volume 3497, pages 659–664. Springer Berlin / Heidelberg, 2005. 1