# A ZGPCA ALGORITHM FOR SUBSPACE ESTIMATION

*Haoran Yi**

Computer Science Department,
University of Illinois at Urbana-Champaign,
IL, USA, 61801

*Deepu Rajan, Liang-Tien Chia*

School of Computer Engineering,
Nanyang Technological University,
Singapore 639798.

**ABSTRACT**

We propose a new algorithm called the ZGPCA algorithm for subspace estimation based on the GPCA (Generalized Principal Component Analysis) algorithm. It is formulated within an FIR filter framework so that the norm vectors of the subspaces correspond to filter coefficients. It is shown that such an approach leads to a more accurate and computationally efficient method compared to the GPCA algorithm. We extend the ZGPCA algorithm to make it recursive so that subspaces with possibly different dimensions can be obtained. We also propose a new distance measure that can be used for $k$-means clustering of sample points within a subspace. Experimental results on synthetic data and applications on face clustering and sports video clustering show good performance of the proposed algorithm.

## 1. INTRODUCTION

Subspace estimation is a useful tool in many applications such as face recognition [1] and object tracking [2]. Several algorithms to estimate the underlying subspace from data have been proposed recently, e.g. K-subspace estimation [3] and EM algorithms for probabilistic principal component analysis [4]. The iterative nature of these algorithms lead to very high computational complexity when the data set is large. Moreover, the performance of these algorithms depend largely on their initialization parameters. Poor initialization could cause the optimization process to get trapped in a local optimum and result in slow convergence.

Vidal et al. [5] propose an algebraic geometric approach to subspace estimation called GPCA. The linear subspaces are represented as a set of polynomials and an iterative polynomial factorization algorithm (PFA) is developed to compute the norms of the subspaces. In [6], they propose another algorithm to compute the norm vectors which are derived by evaluating the polynomial derivatives at samples in the subspaces.

In this paper, we propose a new subspace estimation algorithm, called ZGPCA, based on GPCA by recasting the problem within an FIR filter framework. In doing so, we reduce the problem to determining the filter coefficients, which indeed turn out to be the norm vectors of the subspaces. This is shown to be computationally more efficient and also more accurate in the presence of noise. We extend the FIR filter based ZGPCA algorithm to achieve the same objective as in [6], i.e., to estimate multiple subspaces of unknown and possibly different dimensions. However, it is done in a very simple and intuitive manner using a process of recursion that provides explicit control of the dimensionality of the subspaces to be extracted.

The paper is organized as follows. In section 2, we review the GPCA algorithm. The proposed ZGPCA algorithm is described in section 3. The recursive extension of the ZGPCA algorithm and a new distance measure for $k$-means clustering are described in section 4. The experimental results are presented in section 5 and section 6 presents conclusions.

## 2. REVIEW OF GPCA ALGORITHM

In this section, we review the GPCA algorithm as described in [5]. Consider sample data points $\{\mathbf{x}_j \in R^K\}, j = 1, 2, \ldots, N$, drawn from $m$ $k$-dimensional ($k < K$) linear subspaces of $R^K$, $\{S_i\}, i = 1, \ldots, m$. When the subspace has dimensionality of $k = K - 1$, every $(K - 1)$ dimensional space $S_i$ in $R^K$ can be represented by a nonzero normal vector $\{\mathbf{b}_i\}$ in $R^K$ as $S_i = \{\mathbf{x} \in R^K : \mathbf{b}_i^T\mathbf{x} = 0\}$. Since the subspaces $S_i$ are all distinct from each other, the norm vectors $\{\mathbf{b}_i\}, i = 1, \cdots, m$, are pairwise linearly independent. If every sample point $\mathbf{x}$ in $R^K$ lies on one of the subspaces $S_i$, they satisfy the homogeneous polynomial of degree $m$ in $\mathbf{x}$ with real coefficients $p_m(\mathbf{x}) = \prod_{i=1}^m(\mathbf{b}_i^T\mathbf{x}) = 0$. This nonlinear equation can be linearized according to $p_m(\mathbf{x}) = \nu_{\mathbf{m}}(\mathbf{x})^T\mathbf{c} = \sum C_{n_1,n_2,\cdots,n_K} x_1^{n_1}x_2^{n_2}\cdots x_K^{n_K} = 0$ where $\nu : [x_1, \cdots, x_K]^T \longrightarrow [\cdots, x_1^{n_1}x_2^{n_2}\cdots x_K^{n_K}, \cdots]^T$, is called a Veronese map and $x_1^{n_1}x_2^{n_2}\cdots x_K^{n_K}$ is a monomial with $n_i$'s arranged in degree-lexicographic order. Also, $n_1+\cdots+n_K = m, \ n_j \geq 0, \ j = 1, \cdots, K$. The coefficients $C_{n_1,n_2,\cdots,n_K}$ are functions of the entries in $\{\mathbf{b}_i\}, i = 1, \cdots, m$ and they form the coefficient vector $\mathbf{c}$. The nonlinear Veronese map maps the original data $\{\mathbf{x}_p\}, p = 1, 2, \cdots, N$ of dimension $K$ into an embedded data space of higher dimension $M_m =$

*The first author performed the work while at Nanyang Technological University, Singapore.

$\begin{pmatrix} m+K-1 \\ K-1 \end{pmatrix}$. The problem of GPCA is then to recover $\{\mathbf{b}_i\}$, given the coefficient vector $\mathbf{c}$ of the polynomial $p_m(\mathbf{x})$.

The unknown number of subspaces can be determined from the rank of the Veronese map matrix $L_m$ of the form $[\nu_{\mathbf{m}}(\mathbf{x_1})^T, \nu_{\mathbf{m}}(\mathbf{x_2})^T, \cdots, \nu_{\mathbf{m}}(\mathbf{x_N})^T]^T$. Given the sample points $\mathbf{x_p}$, $\mathbf{c}$ is obtained by solving the set of linear equations [5], $L_m\mathbf{c} = [\nu_{\mathbf{m}}(\mathbf{x}_1)^T \ \nu_{\mathbf{m}}(\mathbf{x}_2)^T \ \cdots \nu_{\mathbf{m}}(\mathbf{x}_N)^T]^T\mathbf{c} = 0$, where $N$ is the number of sample points. The remaining problem is to factorize the polynomial $p_m(\mathbf{x})$ with coefficient vector $\mathbf{c}$ to determine the subspace norm vectors $\{\mathbf{b}_i\}, i = 1, \cdots, m$. The authors describe a polynomial factorization algorithm in which the norm vectors are factorized out from the Veronese map coefficient vector in an iterative manner.

## 3. THE ZGPCA ALGORITHM

The main thrust of the proposed ZGPCA algorithm is on an efficient and accurate computation of the norm vectors $\mathbf{b}_i$ from the Veronese map coefficient vector $\mathbf{c}$.

We note that the number of degrees of freedom (*DoF*) for $\mathbf{c}$ is $\begin{pmatrix} m+K-1 \\ m \end{pmatrix}$, while that for $\mathbf{b}_i$ is $m(K-1)+1$, where $K$ is the dimension of the space and $m$ is the number of subspaces. For a 6D video signal $(x, y, t, r, g, b)$, to extract $m = 3$ subspaces, there are 56 *DoF* for $\mathbf{c}$ and 16 for $\mathbf{b}$. Hence, it is not suitable to compute $\mathbf{b}_i$ from $\mathbf{c}$ directly. We consider $\mathbf{b}_i$ to be FIR filter coefficients that need to be determined. Rewriting $p_m(\mathbf{x})$ as $\prod_{i=1}^{m}(b_{i,0}x_0 + b_{i,1}x_1 + \cdots + b_{i,K-1}x_{K-1}) = 0$ leads to the linear filter system shown in Figure 1. The transfer function $H(z)$ of the filter is obtained as $H(z) = \prod_{i=1}^{m}(b_{i,0} + b_{i,1}z^{-1} + \cdots + b_{i,K-1}z^{-(K-1)})$. If the input $x(n)$ to the filter is a Dirac delta function, the output vector $\mathbf{y}$ is of length $1 + m(K-1)$, which is same as the number of *DoF* for $\mathbf{b}_i$. We now show that $\mathbf{y}$ is related to the Veronese map coefficient vector $\mathbf{c}$ of size $\mathbf{c_m}$, through a sparse transformation matrix $R$, i.e., $\mathbf{y} = R \times \mathbf{c}$. The elements of the matrix denote whether a particular coefficient $C_{n_1,n_2,\cdots,n_K}$ contributes to the output $y(n)$. This can be ascertained through the product $[n_1 \ n_2 \ \ldots \ n_K] \times [0\ 1 \ \ldots \ K-1]^T$, i.e., for every column $j$ in $R$, $R(i,j) = 1$ if $i = [n_1 \ n_2 \ \ldots \ n_K] \times [0\ 1 \ \ldots \ K-1]^T$ and is 0, otherwise.

To illustrate the above idea, consider the example shown in Figure 2 in which two subspaces are represented by the vectors $\mathbf{b}_1 = \frac{1}{\sqrt{62}}[6, -5, 1]^T$, $\mathbf{b}_2 = \frac{1}{\sqrt{3}}[1, -1, 1]^T$. The polynomial is obtained as $p(\mathbf{x}) = \frac{1}{\sqrt{62}}\frac{1}{\sqrt{3}}(6x_1^2 - 11x_1x_2 + 5x_2^2 + 7x_1x_3 - 6x_2x_3 + x_3^2)$. The coefficient vector $\mathbf{c} = [c_{200},\ c_{110},\ c_{020},\ c_{010},\ c_{011},\ c_{002}]^T$ and $R$ is shown in the figure, e.g., for column 3, $i = [0\ 2\ 0] \times [0\ 1\ 2]^T = 2$, and hence $R_{2,2} = 1$.

For delta input, $y(z) = \prod_{i=1}^{m} b_i(z)$ and the zeros of $\mathbf{y}(\mathbf{z})$ and $\mathbf{b_i}$ are the same. The number of zeros for $\mathbf{y}(\mathbf{z})$ is $m(K-1)$ since $y(n)$ is of length $1 + m(K-1)$. Similarly, each arm of the FIR filter $\mathbf{b_i}$ (which corresponds to a subspace) has $K$ taps requiring it to have $K-1$ zeros. Hence, the total

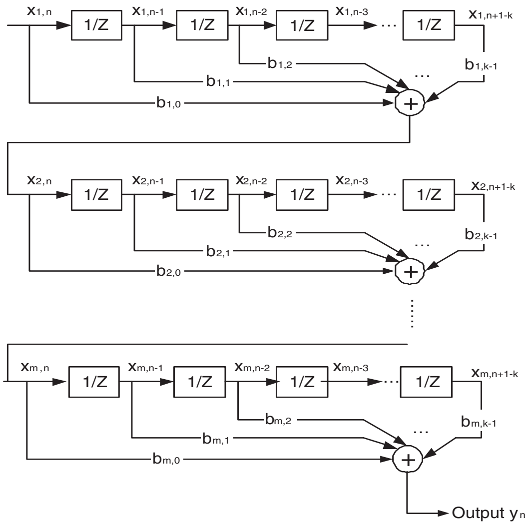


**Fig. 1**. FIR filter formulation for subspace estimation.

number of zeros for the FIR filter with $m$ arms is $m(K-1)$, which is the same as that for $\mathbf{y}(\mathbf{z})$. Therefore, factoring out the subspaces $\mathbf{b}_i$, it is more prudent to use $\mathbf{y}(\mathbf{z})$ rather than the Veronese coefficient map vector $\mathbf{c}$.

Since the FIR filter with $m$ arms shares the zeros with $\mathbf{y}(\mathbf{z})$ exactly, we can solve for the norm vector $\mathbf{b}_i$ by finding an optimal grouping of the zeros of $\mathbf{y}(\mathbf{z})$ into $m$ groups of $(K-1)$ zeros each such that an error function is minimum over all possible groupings. In the current implementation, we use full search to determine the optimal grouping of zeros since the dimension of the space $(K)$ and the number of subspaces $(m)$ are much smaller than the number of sample points $(N)$. The computational cost of grouping zeros is negligible compared to the cost of solving the large linear system, $L_m\mathbf{c} = 0$, for the Veronese coefficient map vector $\mathbf{c}$. In addition to the fact that the transformation of $\mathbf{c}$ into $\mathbf{y}(\mathbf{z})$ facilitates the recovery of $\mathbf{b}_i$ from a space with the same number of free parameters as $\mathbf{b}_i$, it also eliminates the problem of error propagation in the PFA algorithm, thus making the estimates more accurate and reliable.

We return to Figure 2 to illustrate the working of the proposed ZGPCA algorithm in its entirety. Here $K = 3$ and $m = 2$. The transformation of $\mathbf{c}$ with 5 degrees of freedom into $\mathbf{y}$ with 4 degrees of freedom matches $\mathbf{y}$ with $\mathbf{y_1}$ and $\mathbf{y_2}$ together having a total of 4 degrees of freedom. The zeros of $\mathbf{y}(\mathbf{z})$ are grouped into two pairs, viz., $(3, 2), (\frac{1}{2} + \frac{\sqrt{3}}{2}i, \frac{1}{2} - \frac{\sqrt{3}}{2}i)$. The estimate of the norm vector $\mathbf{b'}_1$ corresponding to the pair $(3, 2)$ is obtained from $(z-3)(z-2) = z^2 - 5z + 6$ as [6 -5 1]. Similarly, the estimate of the vector $\mathbf{b'}_2$ is obtained from the conjugate pair of zeros.

$$\text{I.}\quad \mathbf{c} \propto [6 \ \ -11 \ \ 5 \ \ 7 \ \ -6 \ \ 1]^T, \begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ y(3) \\ y(4) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_{200} \\ c_{110} \\ c_{020} \\ c_{101} \\ c_{011} \\ c_{002} \end{bmatrix}$$

$$\text{II.}\quad \mathbf{D} = M \times \mathbf{c} \propto [6 \ \ -11 \ \ 12 \ \ -6 \ \ 1]^T$$

$$\text{III.}\quad \text{Zeros}: (3,2),\ \left(\frac{1}{2} + i\frac{\sqrt{3}}{2}, \frac{1}{2} - i\frac{\sqrt{3}}{2}\right)$$

$$[\mathbf{D}(\mathbf{z}) = 6 - 11z^{-1} + 12z^{-2} - 6z^{-3} + z^{-4}]$$

$$\text{IV.}\quad \mathbf{b}_1' \propto [6 \ \ -5 \ \ 1]^T, \ \ \mathbf{b}_2' \propto [1 \ \ -1 \ \ 1]$$

**Fig. 2**. Example illustration of the ZGPCA algorithm.

## 4. RECURSIVE ZGPCA

The ZGPCA algorithm determines a set of $K-1$ dimension subspaces in $K$ dimension space. In [6], Vidal et al., describe the polynomial differentiation algorithm (PDA) to extract an unknown number of possibly different dimensions of subspaces. We achieve the same objective through a recursive ZGPCA algorithm in which the ZGPCA algorithm is recursively applied to lower dimensional subspaces. The first step is to find $m$ number of $K-1$ dimension subspaces in $K$ dimensions. Subsequently, each of the $K-1$ dimension subspaces is probed for any $K-2$ dimension subspaces using the ZGPCA algorithm. In doing so, we evaluate the error in fitting a $K-2$ dimension subspace using the following cost function: $Err_K(\mathbf{b_1}, \cdots, \mathbf{b_m}) = \frac{1}{m \cdot N} \sum_{i=1}^{m} \sum_{j=1}^{N} \log(\mathbf{b_i}^T \mathbf{x^j})$. The sample points are projected onto the closest $K-1$ dimension subspace and the ZGPCA algorithm is used to estimate a set of $K-2$ dimension subspace embedded in the $K-1$ dimension subspace. If the fitting error $Err_{K-2} - Err_{K-1} \le \alpha$, the $K-2$ dimension subspaces are accepted. $\alpha$ controls the trade off between the fitting error and the dimension of the subspaces such that large $\alpha$ favors more subspaces of lower dimension. In our experiment, $\alpha$ is set to be 1. This recursive decomposition is continued until the fitting error is too large, which implies that the underlying subspaces have been found. The proposed recursive ZPGCA algorithm is simpler and more intuitive than the PDA algorithm described in [6]. Moreover, the dimensionality of the subspaces to be extracted can be controlled by $\alpha$.

A synthetic example to extract subspaces using recursive ZGPCA algorithm is shown in Figure 3. The sample points lie along two coplanar lines in three dimension (Figure 3(a)). The Recursive ZGPCA algorithm estimates a plane in 2D subspaces (Figure 3(b)), which is further decomposed into two lines comprising two 1D subspaces (Figure 3(c)).

### Data clustering

Clustering is an unsupervised classification mechanism that has a variety of applications. In the absence of little or no

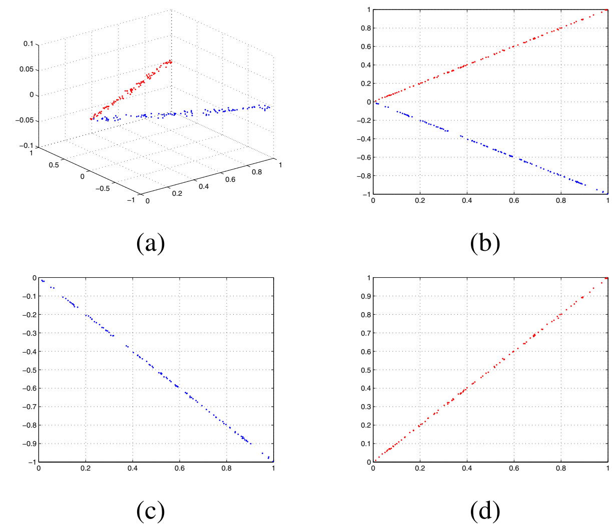


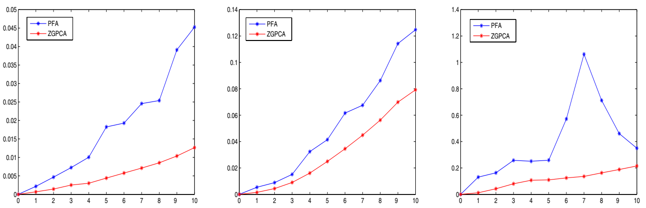
**Fig. 3**. Illustration of RZGPCA algorithm: (a) Two lines in 3D, (b) extracted 2D subspace and (c,d) extracted 1D subspaces.

prior information about the data, an estimate of the underlying subspace, if any, becomes an effective initialization point for subsequent clustering. It is important to include the global configuration of the sample points so that contextual information is also incorporated into the clustering procedure. In this section, we present a new distance measure for data clustering which incorporates such a fusion of local and global information.

After obtaining the set $\mathbf{b}_i$ of subspaces using the proposed recursive ZPGCA algorithm, each sample point's membership in every $\mathbf{b}_i$ is calculated using a membership function $mem_j(\mathbf{b_i}) = \exp\{-d(\mathbf{x}_j, \mathbf{b}_i)\}$ where $d(\mathbf{x}_j, \mathbf{b}_i)$ is the distance of the sample $\mathbf{x}_j$ to the subspace $\mathbf{b}_i$ calculated by projecting $\mathbf{x}_j$ onto $\mathbf{b}_i$. The membership values for $\mathbf{x}_j$ is collected into a membership vector $\mathbf{mem}_j$ and normalized to unit length. The pairwise distance between samples $\mathbf{x}_i$ and $\mathbf{x}_j$ is then defined using a weighted distance function $d(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{mem_i} - \mathbf{mem}_j)^T \cdot W \cdot (\mathbf{mem}_i - \mathbf{mem}_j)$ where $W$ is the weight matrix such that $W_{ij}$ is the principal angle between subspace $i$ and subspace $j$. The pairwise distances are used to perform $k$-means clustering of all the sample points.

## 5. EXPERIMENTAL RESULTS

The synthetc data consists of two dimensional points drawn from $R^3$. We consider $m = 2, 3, 4$ number of subspaces with each subspace containing 200 points. Zero mean Gaussian noise with variance ranging from 1% to 10% of the maximum signal value is added to the sample points. The estimation of the subspaces is carried out 30 times for each noise level.Figure 4 shows the variation of average error in estimation as the noise level increases. We compare the ZGPCA al-




(a) (b) (c)

**Fig. 4**. Comparison of ZGPCA and PFA algorithms in terms of error in norm angles v/s noise level for (a) 2, (b) 3, and (c) 4 subspaces in $R^3$.

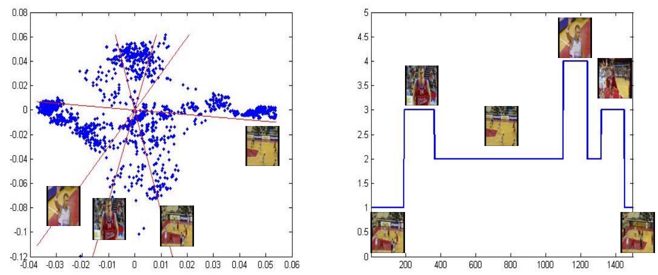
**Table 1**. Classification accuracy (%) of face images.

| Method | F1 | F3 | F5 | F8 | F10 |
|---|---|---|---|---|---|
| PFA | 71.9 | 68.7 | 89.1 | 73.4 | 85.9 |
| ZGPCA | 75.0 | 71.8 | 92.1 | 79.7 | 84.4 |
| ZGPCA+clustering | 89.1 | 81.3 | 100 | 84.4 | 100 |

gorithm with the PFA algorithm of [7]. For $n = 2$ subspaces, the error in estimates at noise level of 10% for the former is about 0.72 degrees while that for the latter is about 2.57 degrees (the y-axis in Figure 4 is in radians). In Figure 4(c), we see that for $n = 3$ subspaces, the error of the PFA algorithm increased dramatically around noise level of 7%, while the error from the proposed ZGPCA algorithm remains consistent.

Next, we consider the problem of clustering faces under varying illumination so that each cluster contains faces of the same person. Since the images of different faces lie on different subspaces [3], we can estimate the basis for each subspace using the proposed ZGPCA algorithm. The face images are chosen from the Yale face database B. There are 320 frontal face images of 5 subjects (1, 3, 5, 8, 10) under 64 different lighting conditions. The face images are rescaled to size $30 \times 40$, and projected on to the first 3 principal components using PCA, since the number of pixels is large compared to the dimension of the subspaces. We extract five subspaces of two dimensions each. We compare the ZGPCA and the PFA algorithms by classifying faces to the nearest subspace found by the two methods. From Table 1 we see that the ZGPCA provides higher classification accruacy than the PFA algorithm for all faces except subject 10. The distance measure is evaluated by clustering the membership values of the sample points in each subspace obtained by the ZGPCA algorithm. The third row in Table 1 shows even higher classification rate and perfect classification for subjects 5 and 10.

The last experiment considers the problem of sports video frame clustering with the objective of obtaining a temporal segmentation of a video sequence. Each frame in a one minute long sequence of 1500 frames is reduced to size $16 \times 16$. The 256 dimension feature vectors are projected into the first two principal component by PCA. Figure 5(a) shows the 2D



(a) (b)

**Fig. 5**. (a) Subspaces extracted from video sequence; (b) Temporal segmentation with representative frames.

plot of the estimated subspaces with corresponding exemplar frames. The temporal segmentation of the sequence is shown in Figure 5(b) and the representative frame (the one nearest to the cluster center) from each segment is also shown.

## 6. CONCLUSIONS

In this paper, we have proposed an FIR filter based subspace estimation algorithm based on the GPCA algorithm in which the filter coefficients are the norm vectors of the subspaces. Such a framework results in a more accurate and robust estimation of subspaces than the GPCA algorithm. Moreover, the subspaces are estimated directly rather than in an iterative fashion. We also propose a recursive form of the algorithm to estimate subspaces with possibly different dimensions. The experimental results demonstrate the utility of the proposed algorithms.

## 7. REFERENCES

[1] Xiaoou Tang and Zhifeng Li, "Frame synchronization and multi-level subspace analysis for video based face recognition," in *Proc. of IEEE CVPR*, June 2004, vol. 2, pp. 902–907.

[2] K. Kanatani, "Motion segmentation by subspace separation and model selection," in *Proc. of IEEE ICCV*, 2001, vol. 2, pp. 586 – 591.

[3] J. Ho, M.-H. Yang, J. Lim, K.-C. Lee, and D. Kriegman, "Clustering apperances of objects under varying illumination conditions," in *Proc. of IEEE CVPR*, 2003, pp. 11–18.

[4] T. Su and J. Dy, "Automated hierarchical mixtures of probabilistic principal component analyzers," in *Proc. of ICML*, Banff, Canada, July 2004.

[5] Rene Vidal, Yi Ma, and Shankar Sastry, "Generalized principal component analysis (GPCA)," *IEEE Transactions on PAMI*, vol. 27, no. 12, pp. 1945–1959, 2005.

[6] R. Vidal, Y. Ma, and J. Piazzi, "A new GPCA algorithm for clustering subspaces by fitting, differentiating and dividing polynomials," in *Proc. of IEEE CVPR*, 2004, vol. 1, pp. 510–517.

[7] R. Vidal, Y. Ma, and S. Sastry, "Generalized principle component analysis (GPCA)," in *Proc. of IEEE CVPR*, June 2003, vol. 1, pp. 621–628.