SPARSE REPRESENTATION OF IMAGES WITH HYBRID LINEAR MODELS

Kun Huang *Allen Y. Yang* *Yi Ma*
Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801
Email: {kunhuang, yangyang, yima}@uiuc.edu

ABSTRACT

We propose a mixture of multiple linear models, also known as hybrid linear model, for a sparse representation of an image. This is a generalization of the conventional Karhunen-Loeve transform (KLT) or principal component analysis (PCA). We provide an algebraic algorithm based on generalized principal component analysis (GPCA) that gives a global and non-iterative solution to the identification of a hybrid linear model for any given image. We demonstrate the efficiency of the proposed hybrid linear model by experiments and comparison with other transforms such as the KLT, DCT, and wavelet transforms. Such an efficient representation can be very useful for later stages of image processing, especially in applications such as image segmentation and image compression.

1. INTRODUCTION

In image processing, one often seeks a more efficient representation for a digital image than the pixel-based representation. A typical approach is to divide the image into a set of blocks. If there is strong statistical correlation between the blocks, they can be represented as the superposition of a much smaller number of components, also known as *the sparse components* [17]. Traditional approaches to representing an image by a different basis include using the discrete Fourier transform (DFT) or the discrete cosine transform (DCT). A major drawback for these transforms is that, despite significantly different statistics in different images, the bases of these transforms are fixed. So there are reasons to believe that a more efficient representation can be achieved with the so-called *adaptive basis* [20]. One method to obtain such an adaptive basis is via the Karhunen-Loeve transform (KLT), also known as principal component analysis (PCA) in machine learning. If a set of (random) vectors obey a common second-order statistical model, the basis (or the linear model) identified by the KLT is in fact *optimal* [10, 23]. But it will lose its optimality if a single image contains regions with significantly different textures that *cannot* be described by a single linear statistical model, which is, unfortunately, typically the case for a generic image.

This brings up the fundamental problem that this paper is about to address: How can we simultaneously *segment* the image into different regions and *estimate* an adaptive basis for each region? That is, to fit a mixture of linear models, also known as *hybrid linear model*, to the image, but without knowing *a priori* how many linear models to use, the dimension of each model, or which model applies to which blocks.

In the machine learning literature, the problem of simultaneously estimating a mixture of models and segmenting data into respective models was usually resolved via an incremental scheme that *iterates* between segmentation and estimation, e.g., the expectation maximization (EM) method. It has only recently been discovered that a non-iterative and global solution, called generalized principal component analysis (GPCA) [24, 13], exists for the segmentation and estimation of hybrid linear models. The idea that image segmentation may improve image compression is *not* new. However, we believe that GPCA is a method that can seamlessly combine these two key components in image processing. It offers the new capability to represent different image regions with different colors and textures by different linear models with different linear bases.

Relation to prior work. There is a vast amount of literature on finding adaptive bases (or transforms) for signals. Adaptive wavelet transforms and adapted wavelet packets have been extensively studied [4, 20, 15, 6, 18]. The idea is to search for an optimal transform from a limited (although large) set of possible transforms. Another approach is to find some universal optimal transform based on the signals [10, 19, 6]. Spatially adapted bases have also been developed such as [2, 21, 16]. The main purpose of this paper is to show that an adaptive basis based on a single model (such as the KLT and PCA) is not necessarily optimal for an efficient image representation, and hybrid models may be a much better choice.

The notion of a mixture of linear models and bases for image representation is closely related to the *sparse component analysis* [17]. That is to identify a set of non-orthogonal base vectors for natural images such that the representation of the images is sparse. In the related work of [3, 11, 12, 8, 22], the main goal is to find a mixture of models such that the signals can be decomposed into multiple models and their overall representation is sparse. In that approach, the signals are expressed as a linear superposition of all the models while, in this paper, the signals will be segmented to mutually exclusive groups, and a sparse representation is found for each group.

Image segmentation based on local color and texture information extracted from various filter banks has been studied extensively in the computer vision literature (e.g., [1, 14, 7]). Since the MPEG-4 have started to incorporate texture segmentation [9], we expect that the concept and method introduced in this paper will be useful for developing new image and video processing techniques.

2. REPRESENTATION OF AN IMAGE WITH A HYBRID LINEAR MODEL

In this section, we introduce the notion of hybrid linear models for images. Normally, we divide a digital image

0-7803-8554-3/04/$20.00 ©2004 IEEE.

I into a set of, say N, non-overlapping[1] equal-size $l \times l$ blocks, that is, $I = \cup_{j=1}^{N} B_i$. Denote the number of color channels of the image to be c. For grayscale images $c = 1$, and for color images $c = 3$ (i.e., RGB, HSV, or YCbCr). Then we may represent each block B by a vector $\boldsymbol{x}$ that collects the pixel values, and the dimension of the vector $\boldsymbol{x}$ is $K = cl^2$.

As we have contended in the introduction, a single linear model has its limitations when applied to a generic image which often consists of regions with significantly different textures. However, it is reasonable to assume that a linear model is still valid at least for each region, if we know how to segment the image into such regions. That is, we assume that the image blocks can be segmented into multiple groups: $\boldsymbol{X} = \cup_{i=1}^{n} \boldsymbol{X}_i$, and for each group $\boldsymbol{X}_i$ there exists a basis $\boldsymbol{B}_i = \{\boldsymbol{b}_{ij}\}_{j=1}^{k_i}$ such that $\boldsymbol{x} = \sum_{j=1}^{k_i} \alpha_j \boldsymbol{b}_{ij}$, if $\boldsymbol{x} \in \boldsymbol{X}_i$. We denote the subspace spanned by the basis $\boldsymbol{B}_i$ by $S_i \doteq \text{span}(\boldsymbol{B}_i)$, $i = 1, 2, \ldots, n$. Let k_i be the dimension of S_i and k the dimension of span$(\cup_{i=1}^{n} S_i)$. Then the *average number of components* per block needed for the hybrid representation, $\frac{1}{N}(k_1 N_1 + \cdots + k_n N_n)$,[2] would in general be much less than the k components if a single linear model is used instead.[3] Such a hybrid linear model can be intuitively illustrated by Fig. 1.

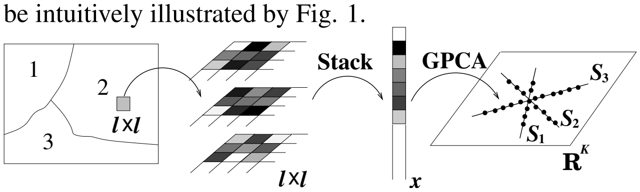

Fig. 1. The representation of an $l \times l$ color image block. The pixels with three color channels are first stacked as a single vector $\boldsymbol{x}$ of dimension $K = 3l^2$. The hybrid linear model assumes that all these vectors lie in the union of multiple subspaces $S_1, S_2, \ldots, S_n$ of dimension $k_1, k_2, \ldots, k_n$, respectively, in $\mathbb{R}^K$.

There are several difficulties in identifying such a hybrid linear model for an image: 1. We do not know the number of linear subspaces n; 2. We do not know the basis $\boldsymbol{B}_i$ or even the dimension k_i of each linear subspace; 3. We do not know which blocks $\boldsymbol{x}$ belong to which group $\boldsymbol{X}_i$; and 4. The subspaces spanned by different bases might have nontrivial intersection. Fortunately, these are exactly the type of problems that GPCA is designed to solve [24, 13].

3. IDENTIFYING HYBRID LINEAR MODELS VIA GPCA

In this section, we will give a brief introduction to the estimation of a hybrid linear model, and outline the essential ideas of GPCA. The interested readers may refer to [24] and [13] for all the technical details.

[1]The non-overlapping assumption can be relaxed for pixel-wise image segmentation.

[2]Here $N_i = |\boldsymbol{X}_i|$ is the number of elements in the group $\boldsymbol{X}_i$.

[3]The overhead of book keeping the extra base vectors and the membership of each group is negligible as long as N is very large and n is relatively small.

A subspace $S_i \subset \mathbb{R}^K$ of dimension k_i, where $0 < k_i < K$, can be represented with $K - k_i$ polynomials of degree *one* (linear equations) of the form $S_i = \{\boldsymbol{x} \in \mathbb{R}^K : \boldsymbol{C}_i^T \boldsymbol{x} = 0 \Leftrightarrow \bigwedge_{j=1}^{K-k_i} (\boldsymbol{c}_{ij}^T \boldsymbol{x} = 0)\}$ where $\boldsymbol{C}_i \doteq [\boldsymbol{c}_{i1}, \ldots, \boldsymbol{c}_{i(K-k_i)}] \in \mathbb{R}^{K \times (K-k_i)}$ is a basis for the orthogonal complement of S_i, $S_i^\perp$. Similarly, a mixture of n subspaces $\{S_i \subset \mathbb{R}^K\}_{i=1}^n$ can also be represented as the zeros of a set of homogeneous polynomials. That is, $\bigvee_{i=1}^n (\boldsymbol{x} \in S_i) \Leftrightarrow \bigvee_{i=1}^n \bigwedge_{j=1}^{K-k_i} (\boldsymbol{c}_{ij}^T \boldsymbol{x} = 0) \Leftrightarrow \bigwedge_\sigma \bigvee_{i=1}^n (\boldsymbol{c}_{i\sigma(i)}^T \boldsymbol{x} = 0)$, where the right hand side (RHS) is obtained by exchanging ands and ors using De Morgan's laws, and σ represents a particular choice of one normal vector $\boldsymbol{c}_{i\sigma(i)}$ from each basis $\boldsymbol{C}_i$. Notice that each one of the $\prod_{i=1}^n (K - k_i)$ equations in the RHS is of the form

$$\bigvee_{i=1}^{n} (\boldsymbol{c}_{i\,\sigma(i)}^T \boldsymbol{x} = 0) \Leftrightarrow \prod_{i=1}^{n} (\boldsymbol{c}_{i\,\sigma(i)}^T \boldsymbol{x}) \doteq p_{n\sigma}(\boldsymbol{x}) = 0, \quad (1)$$

which is simply a homogeneous polynomial of degree n in K variables that is factorizable into a product of n linear forms in $\boldsymbol{x}$. There are in total $M_n = \begin{pmatrix} n+K-1 \\ n \end{pmatrix}$ monomials of degree n in K variables. Therefore, we can write each one of those polynomials as a linear combination of a vector of coefficients $\beta_n \in \mathbb{R}^{M_n}$ as

$$p_n(\boldsymbol{x}) = \beta_n^T \nu_n(\boldsymbol{x}) = \sum_{n_1+\cdots+n_K=n} \beta_{n_1,\ldots,n_K} x_1^{n_1} \cdots x_K^{n_K}, \quad (2)$$

where $\nu_n(\boldsymbol{x}) \in \mathbb{R}^{M_n}$ is the Veronese embedding of degree n.

Then we have the following key results from the GPCA method [24].

Theorem 1 (Algebraic GPCA) *A collection of n subspaces can be described as the set of points satisfying a set of homogeneous polynomials of the form $p(\boldsymbol{x}) = \prod_{i=1}^n (\boldsymbol{c}_i^T \boldsymbol{x}) = \beta^T \nu_n(\boldsymbol{x}) = 0$, where $\boldsymbol{c}_i \in \mathbb{R}^K$ is a normal vector to the ith subspace S_i. When n is known and a sufficient number of points $\{\boldsymbol{x}_i\}_{i=1}^N$ are given on the subspaces, one can estimate all such polynomials from the null-space of the embedded data matrix $L_n = [\nu_n(\boldsymbol{x}_1), \ldots, \nu_n(\boldsymbol{x}_N)]^T$, and the normal vectors $\{\boldsymbol{c}_i\}$ to the ith subspace from the derivative of the polynomials $\{p(\boldsymbol{x})\}$, $\{Dp(\boldsymbol{x})\}$, at a point $\boldsymbol{x} = \boldsymbol{x}_i$ in the ith subspace. The dimension of the ith subspace is obtained as $k_i = K - rank(span_p Dp(\boldsymbol{x}_i))$.*

Example 1 *Imagine we are given data in $\mathbb{R}^3$ drawn from the line $S_1 = \{\boldsymbol{x} : x_1 = x_2 = 0\}$ and the plane $S_2 = \{\boldsymbol{x} : x_3 = 0\}$, as shown in Fig. 2. We can describe the entire data set as: $S_1 \cup S_2 = \{\boldsymbol{x} : (x_1 = x_2 = 0) \vee (x_3 = 0)\} = \{\boldsymbol{x} : (x_1 x_3 = 0) \wedge (x_2 x_3 = 0)\}$. Therefore, the mixture of subspaces is in this case described by the polynomials $p_1(\boldsymbol{x}) = x_1 x_3$ and $p_2(\boldsymbol{x}) = x_2 x_3$, both of degree 2. Let $P(\boldsymbol{x}) = [p_1(\boldsymbol{x}),\ p_2(\boldsymbol{x})]$. Then the derivatives of $P(\boldsymbol{x})$ at points $\boldsymbol{y}_1 = [0, 0, 1]^T \in S_1$ and $\boldsymbol{y}_2 = [1, 1, 0]^T \in S_2$ are:*

$$DP(\boldsymbol{x}) = \begin{bmatrix} x_3 & 0 \\ 0 & x_3 \\ x_1 & x_2 \end{bmatrix} \Rightarrow DP(\boldsymbol{y}_1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix},\ DP(\boldsymbol{y}_2) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}.$$

Then the columns of $DP(\boldsymbol{y}_1)$ span $S_1^\perp$ and the columns of $DP(\boldsymbol{y}_2)$ span $S_2^\perp$ (see Fig. 2). Subsequently, the dimension of the line can be retrieved as $k_1 = 3 - rank(DP(\boldsymbol{y}_1)) = 1$, and the dimension of the plane is $k_2 = 3 - rank(DP(\boldsymbol{y}_2)) = 2$.

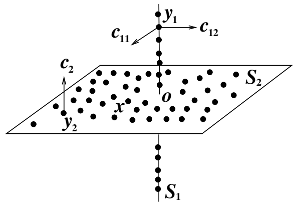

Fig. 2. Data samples drawn from a mixture of one plane and one line in $\mathbb{R}^3$. Arrows are normal vectors.

The above theorem suggests a very simple and effective scheme for estimating a hybrid linear model from the data set. However, firstly, since the number n of subspaces (or linear models) is also unknown, we do not know the correct degree of the Veronese embedding. Secondly, the real data set from the experiments will not be noise free, hence, like in the KLT and PCA, the linear algebraic model should be chosen as the one with the highest fidelity to the noisy data. Fortunately, these two problems can be resolved by a *recursive and robust* version of the above algebraic GPCA algorithm [13]. It can both tolerate noise in the data and handle outliers.

4. IMPLEMENTATION AND EXPERIMENTAL VERIFICATION

In principle, one may use the GPCA method to identify a hybrid linear model for an image. However, a difficulty associated with directly applying the GPCA method to the block vectors is that their dimension K is very large. Even for 8×8 blocks of a grayscale image, $K = 64$ is already too large for the Veronese map to be implemented on a typical computer. Therefore, before applying GPCA, we should reduce the dimensionality by projecting the data onto a lower-dimensional subspace. Theoretically, when the largest dimension of the subspaces, denoted by k_{max}, is known, one may choose a $(k_{max}+1)$-dimensional (affine) subspace $\mathcal{P}$ such that, by projecting $\mathbb{R}^K$ onto this subspace:

$$\pi_{\mathcal{P}} : \; \boldsymbol{x} \in \mathbb{R}^K \quad \mapsto \quad \boldsymbol{x}' = \pi_{\mathcal{P}}(\boldsymbol{x}) \in \mathcal{P}.$$

The resulting $\mathcal{P}$ gives a “minimum” representation that preserves the multilinear structure of the original data.

For the rest of the experiments, we first apply PCA to reduce the dimension of the vectors and then perform GPCA on their (projected) new coordinates.[4] The results from GPCA will be a mixture of linear subspaces and each subspace contains a group of image blocks expressed as the superposition of its basis.

4.1. Hybrid linear model and segmentation

Fig. 3 shows the result of applying the GPCA method to the standard 512×512 Barbara image. The total number of blocks is $N = 4096$. These blocks are segmented automatically by GPCA into three groups as shown in Fig. 3. We can further improve the basis of each subspace by applying PCA to only the image blocks of each group. Fig. 4 displays

Fig. 3. The segmentation of the 4096 image blocks from the Barbara image. The image (left) is segmented into three groups (right three). Roughly speaking, the first group contains mostly image blocks with homogeneous textures; the second and third groups contain blocks with textures of different spatial orientations and frequencies.

the three bases for the three groups shown in Fig. 3, respectively. It is worth noticing that these bases are very consistent with the textures of the image blocks in the respective groups. Therefore, the proposed hybrid linear model can be

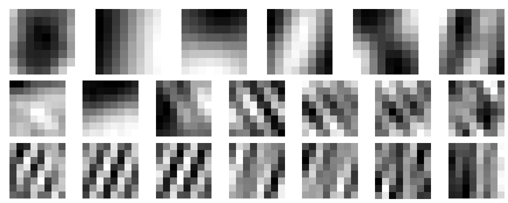

Fig. 4. The three bases for the three groups in Fig. 3, respectively. The number of base vectors for each group is determined by the GPCA method as the dimension of each linear model.

easily extended to pixel-wise texture-based image segmentation. The interested reader may refer to [13].

4.2. Image compression

In addition to image segmentation, the hybrid linear model offers a very efficient representation of the image that is useful for later stages of image processing, especially image compression. In this section, we compare GPCA with other popular transforms for image compression. These transforms are the DCT, KLT (or PCA), and the wavelet transforms. For simplicity and clarity, the comparison does not include any quantization or entropy encoding of the coefficients of each transform.[5] The results clearly demonstrate the superior efficiency of GPCA over any single-model transforms (i.e., the fixed-basis DCT and the adaptive KLT), as well as reveal the relative position of GPCA against the multi-resolution wavelet transforms.[6]

We use the Baboon color image (Fig. 5 left). For the wavelet transforms, we select two filters: one is the simplest, the level-2 Haar transform, and the other is more commonly used (e.g., in JPEG-2000), the level-4 bior4.4 transform [5]. Both are available in the Matlab wavelet toolbox. For all the transforms, we keep 24 nonzero components per 8×8 image block (that is, 8 per color channel).[7] The result is shown in Fig. 5.

[4] In our current implementation, due to the hardware limitation, the dimension for the initial PCA is chosen to be ≤ 12. Furthermore, we can only sample 2000 blocks from the whole image to compute the hybrid model and then project the rest blocks onto their closest subspaces.

[5] For such topics, we refer the reader to the literature, e.g., [10].

[6] Note that the wavelet transforms are not in the class of transformations with a fixed block-size. Therefore, the comparison would not be fair.

[7] For GPCA, 24 is the upper bound for the average number of components per block, see Section 2.

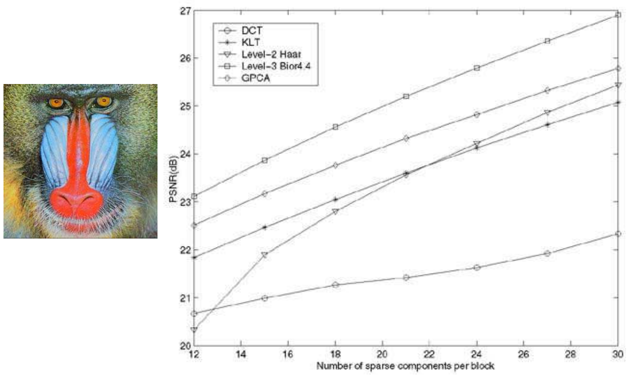

Fig. 5. The PSNR versus the number of components for the DCT, KLT, wavelets (level-2 Haar and level-4 bior4.4), and GPCA for approximating the Baboon image (shown on the left).

From the figure, we note that, the GPCA-based hybrid linear model increases the PSNR of the KLT by about 1dB, but the bior4.4 wavelet has a PSNR about 1dB higher than GPCA. The difference is rather consistent when the number of components is increased. We have also conducted similar experiments on many other classical grayscale and color images, and the results are similar. As conclusion, the GPCA-based hybrid linear model is indeed more efficient than the supposedly optimal KLT (or PCA) among transforms with a fixed block-size; its performance is actually comparable to that of a moderate multi-resolution wavelet transform.

5. CONCLUSIONS

In this paper, we contend that hybrid linear models offer a more powerful modeling paradigm for a sparse representation of images than the conventional single linear models (i.e., the KLT and PCA); we have shown that there is a global and non-iterative solution to the identification of hybrid linear models based on simple linear and polynomial algebraic techniques; we have demonstrated by experiments and comparison with other representations such as the DCT, KLT, and wavelet transforms the efficiency of the hybrid linear representations of images. Therefore, we believe that the hybrid linear models can provide efficient sparse representations of images for later stages of image processing, especially in applications such as image segmentation and compression.

6. REFERENCES

[1] S. Belongie, C. Carson, H. Greenspan, and J. Malik. "Color- and texture-based image segmentation using EM and its application to content-based image retrieval," *ICCV*, pages 675–682, 1998.

[2] J.-Q. Chen, T. N. Pappas, A. Mojsilovic, and B. E. Rogowitz. "Image segmentation by spatially adaptive color and texture features," *ICIP*, 2003.

[3] S.S. Chen, D.L. Donoho, and M.A. Saunders. "Atomic decomposition by basis pursuit.," *SIAM Journal of Scientific Computing*, 20(1):33–61, 1998.

[4] R.R. Coifman and M.V. Wickerhauser. "Entropy-based algorithms for best bases selection," *IEEE Trans. on Information Theory*, 38(2):713–718, 1992.

[5] I. Daubechies. "Orthogonal bases of compactly supported wavelets," *Communications in Pure and Applied Mathematics*, 41:909–996, 1988.

[6] P. Delsarte, B. Macq, and D.T.M. Slock. "Signal-adapted multiresolution transform for image coding," *IEEE Trans. on Information Theory*, 38:897–903, 1992.

[7] Y. Deng and B.S. Manjunath. "Unsupervised segmentation of color-texture regions in images and video," *IEEE Trans. PAMI*, 23(8):800–810, 2001.

[8] D.L. Donoho and M. Elad. "Optimally sparse representation in general (nonorthogonal dictionaries via L^1 minimization," *Proceedings of National Academy of Sciences*, 100(5):2197–2202, 2003.

[9] T. Ebrahimi. "MPEG-4 video verification model version 10.0," *Technical Report, ISO/IEC JTC1/SC29/WG11*, 1998.

[10] M. Effros and P.A. Chou. "Weighted universal transform coding: Universal image compression with the Karhunen-Loeve transform," *ICIP*, volume II, pages 61–64, 1995.

[11] M. Elad and A.M. Bruckstein. "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. on Information Theory*, 48(9):2558–2567, 2002.

[12] A. Feuer and A. Nemirovski. "On sparse representation in pairs of bases," *IEEE Trans. on Information Theory*, 49(6):1579–1581, 2003.

[13] K. Huang, Y. Ma, and R. Vidal. "Minimum effective dimensions for mixtures of subspaces: A robust GPCA algorithm and its applications, submitted to," *CVPR*, 2004.

[14] Z. Kato and T.-C. Pong. "A Markov random field image segmentation model using combined color and texture features," *CAIP*, pages 547–554. Springer-Verlag, 2001.

[15] F.G. Meyer. "Fast adaptive wavelet packet image compression," *IEEE Trans. on Image Processing*, 9(5):792–800, 2000.

[16] D.D. Muresan and T.W. Parks. "Adaptive principal components and image denoising," *ICIP*, 2003.

[17] B.A. Olshausen and D.J.Field. "Wavelet-like receptive fields emerge from a network that learns sparse codes for natural images," *Nature*, 1996.

[18] V. Pavlovic, P. Moulin, and K. Ramchandran. "An integrated framework for adaptive subband image coding," *IEEE Trans. on Signal Processing*, 1998.

[19] H.R. Rabiee, R.L. Kashyap, and S.R. Safavian. "Adaptive multiresolution image coding with matching and basis pursuits," *ICIP*, 1996.

[20] K. Ramchandran, M. Vetterli, and C. Herley. "Wavelets, subband coding, and best basis," *Proceedings of the IEEE*, 84(4):541–560, 1996.

[21] T. Sikora and B. Makai. "Shape-adaptive DCT for generic coding of video," *IEEE Trans. Circuits Syst. Video Technol.*, 5:59–62, 1995.

[22] J.-L. Starck, M. Elad, and D.L. Donoho. "Image decomposition: Separation of texture from piecewise smooth content," In *SPIE*, 2003.

[23] H. Stark and J.W. Woods. *Probability and Random Processes with Applications to Signal Processing*. Prentice Hall, 3rd edition, 2001.

[24] R. Vidal, Y. Ma, and S. Sastry. "Generalized principal component analysis (GPCA): Subspace clustering by polynomial factorization, differentiation, and division," Technical Report UCB/ERL, UC Berkeley, August 15 2003.