Eigendecomposition of Correlated Images Characterized by Three Parameters*

Kishor Saitwal and Anthony A. Maciejewski
Dept. of Electrical and Computer Eng.
Colorado State University
Fort Collins, CO 80523-1373, USA
Email: {Kishor.Saitwal, aam}@colostate.edu

Rodney G. Roberts
Dept. of Electrical and Computer Eng.
Florida A & M - Florida State University
Tallahassee, FL 32310-6046, USA
Email: rroberts@eng.fsu.edu

Abstract

Most eigendecomposition algorithms operate on correlated images that are characterized by only one parameter. Hence they lack the required specifications of fully general 3D image data sets, in which the images need to be characterized by three parameters. In this paper, an extension of one of the fastest known eigendecomposition algorithms is successfully implemented to improve the computational efficiency of computing the eigendecomposition of such 3D image sets. This algorithm can be used in pattern recognition applications such as fully general 3D pose estimation of objects.

1 Introduction

Eigendecomposition-based techniques play an important role in numerous image processing and computer vision applications. The advantage of these techniques is that they are purely appearance based and require few online computations. Singular value decomposition (SVD) is one such eigendecomposition technique that has been used extensively in a variety of applications including pattern recognition [1–3]. All of these applications take advantage of the fact that a set of highly correlated images can be approximately represented by a small set of eigenimages. However, the offline calculation required to determine both the appropriate number of eigenimages as well as the eigenimages themselves can be prohibitively expensive. In particular, common SVD algorithms to compute the complete SVD of a general $m \times n$ matrix require on the order of mn^2 flops. The computational complexity of the SVD has been the subject of much research [4–10] with one of the fastest known algorithms described in [10].

Most previous eigenimage calculations considered images that varied in only one dimension, typically time. These image sets will be referred to as 1D image sets. However, certain pattern recognition applications require that different views of an object taken from different 3D spatial camera locations be considered in the training data set. Because they are characterized by three parameters, such image data sets will be referred to as 3D image sets. The goal of this paper is to extend Chang's algorithm [10] to effectively compute the SVD of 3D image sets.

This paper is organized as follows. Section 2 gives the fundamentals of applying the SVD to 1D image sets and a brief overview of Chang's SVD algorithm. Section 3 explains the generation of fully general 3D image sets, while Section 4 explains how Chang's algorithm can be extended to compute the SVD of such image sets. Supporting results are shown in Section 5, while Section 6 concludes this paper.

2 Eigendecomposition of 1D image sets

Consider an image set containing n correlated images (intensity normalized between 0 and 1) with m pixels each ($m \gg n$). The images are row-scanned to obtain the corresponding image vectors $\mathbf{x}_i$ and the $m \times n$ image data matrix $X = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]$ is generated. The SVD of X is given by

$$X = U\Sigma V^T \tag{1}$$

*This work was supported by the National Imagery and Mapping Agency under contract no. NMA201-00-1-1003 and through collaborative participation in the Robotics Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0012. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

where $U \in \Re^{m\times m}$ and $V \in \Re^{n\times n}$ are orthogonal, and $\Sigma = [\Sigma_d \ \mathbf{0}]^T \in \Re^{m\times n}$ where Σ_d = diag$(\sigma_1, \cdots, \sigma_n)$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ and $\mathbf{0}$ is an n by $m-n$ zero matrix. The columns of U, denoted $\mathbf{u}_i$, $i = 1, \cdots, m$, are referred to as the eigenimages of X; the columns of V, denoted $\mathbf{v}_i$, $i = 1, \cdots, n$, are referred to as the right singular vectors of X, while the diagonal entries σ_i of Σ give the corresponding singular values of X.

In practice, true eigenimages $\mathbf{u}_i$ are not computed exactly, and instead estimates $\mathbf{e}_1, \cdots, \mathbf{e}_k$ that form a k-dimensional basis are used. To quantify the accuracy of these estimates, the measure "energy recovery ratio," denoted ρ, is used here, which is given by

$$\rho(X, \mathbf{e}_1, \cdots, \mathbf{e}_k) = \frac{\sum_{i=1}^{k} \|\mathbf{e}_i^T X\|_2^2}{\|X\|_F^2} \tag{2}$$

where $\|\cdot\|_F$ denotes the Frobenious norm. Note that the true eigenimages give the optimal energy recovery ratio, i.e., $\mathbf{e}_i = \mathbf{u}_i$ with $k = n$ gives $\rho = 1$.

In general, the intensity values of pixels in the correlated images in X vary slowly over time. Chang et al. [10] used this property to show that the right singular vectors $\mathbf{v}_i$ of X are spanned by a few low-frequency sinusoids and the dominant frequencies in their power spectra increased approximately linearly with their index. Consequently, projection of the row space of X (i.e., variation in each pixel) to a smaller subspace spanned by a few low-frequency sinusoids can be used to improve the computational efficiency of the SVD of X. In particular, consider a single sinusoid with a frequency of α, denoted $\mathbf{f}_\alpha$, elements of which are given by

$$\mathbf{f}_\alpha(x) = \frac{1}{\sqrt{n}} e^{-j2\pi\alpha x/n} \tag{3}$$

with $0 \leq x \leq n-1$. Then the orthonormal basis that consists of sinusoids with increasing frequencies can be found using the basis of the 1D discrete Fourier transform (DFT) and the corresponding *real* Fourier matrix H given by

$$\begin{aligned} H &= \left[\begin{array}{cccccc} \mathbf{h_0} & \mathbf{h_1} & \mathbf{h}_2 & \mathbf{h_3} & \mathbf{h}_4 & \cdots \end{array}\right] \\ &= \sqrt{2}\left[\begin{array}{cccccc} \frac{1}{\sqrt{2}}\mathbf{f}_0 & \Re\mathbf{f}_1 & \Im\mathbf{f}_1 & \Re\mathbf{f}_2 & \Im\mathbf{f}_2 & \cdots \end{array}\right] \\ &= \sqrt{\frac{2}{n}}\left[\begin{array}{cccccc} \frac{1}{\sqrt{2}} & c_0 & -s_0 & c_0 & -s_0 & \cdots \\ \frac{1}{\sqrt{2}} & c_1 & -s_1 & c_2 & -s_2 & \cdots \\ \frac{1}{\sqrt{2}} & c_2 & -s_2 & c_4 & -s_4 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots & \cdots \\ \frac{1}{\sqrt{2}} & c_n & -s_n & c_{2n} & -s_{2n} & \cdots \end{array}\right] \end{aligned} \tag{4}$$

where $c_k = \cos(\frac{2\pi k}{n})$ and $s_k = \sin(\frac{2\pi k}{n})$, while $\Re$ and $\Im$ denote the real and the imaginary part, respectively. Chang et al. showed that if p is such that the power spectra of the first k $\mathbf{v}_i$'s of X are restricted to the band $[0, 2\pi p/n]$, then the first k eigenimages $\mathbf{e}_i$ of XH_p are a good approximation to those of X, where H_p denotes the matrix containing the first p columns of H.

Now consider the frequency domain representation of a single row of X (denoted $\mathbf{g}$), which can be computed using the 1D DFT as follows:

$$\begin{aligned} \mathbf{G}(\alpha) &= \frac{1}{\sqrt{n}} \sum_{x=0}^{n-1} \mathbf{g}(x) e^{-j2\pi\alpha x/n} \\ &= \sum_{x=0}^{n-1} \mathbf{g}(x)\mathbf{f}_\alpha(x). \end{aligned} \tag{5}$$

Then (3), (4), and (5) clearly indicate that the multiplication XH can be alternately performed using fast Fourier transform (FFT) techniques for better computational efficiency [10].

3 Generating 3D Image Sets

Fig. 1 illustrates the setup for generating fully general 3D image sets, in which the correlated images are characterized by three parameters instead of one. In this setup, camera locations are defined in a spherical patch above the object with two consecutive camera locations separated by an equiangular distance in that patch. The range of these camera locations is characterized by two parameters, i.e., α_l and β_m, while the third parameter γ_n characterizes image plane rotation to capture different views of the object in equal increments. In practice, the required images can be captured using a video camera attached to a robot end-effector. The robot movement can be controlled to position the camera in one of the specified locations in the spherical patch and then the robot end-effector can be rotated to rotate the image plane of the camera for capturing different orientations of the object from the same location.

The image data matrix for the resulting 3D image sets is given by

$$\begin{aligned} X_3 = [&\mathbf{x}_{111}, \cdots, \mathbf{x}_{L11}, \mathbf{x}_{121}, \cdots \mathbf{x}_{L21}, \cdots, \mathbf{x}_{LM1}, \\ &\mathbf{x}_{112}, \cdots, \mathbf{x}_{LM2}, \cdots, \mathbf{x}_{LMN}] \end{aligned} \tag{6}$$

where an image vector $\mathbf{x}_{lmn}$ corresponds to the row-scanned image of an object taken from camera location (l, m) at image plane rotation n, where $1 \leq l \leq L$, $1 \leq m \leq M$, and $1 \leq n \leq N$.

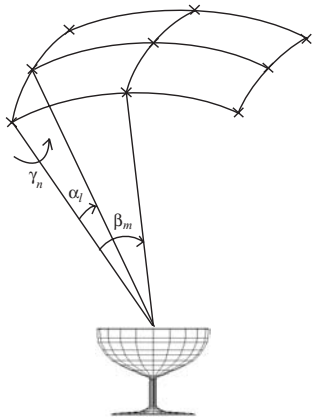

Figure 1. This figure shows the simulation setup for generating 3D image sets, in which the images are characterized by three parameters, i.e., α_l, β_m, and γ_n. The crosses (x) denote the simulated camera locations that are placed in the spherical patch above the object. The range of the parameters α_l and β_m is from -45^o to 45^o with respect to nadir, whereas γ_n ranges from 0 to 360 degrees.

4 Extension of Chang's SVD Algorithm to 3D Image Sets

Consider a three-dimensional signal $g(x, y, z)$ containing L, M, and N samples in x, y, and z dimensions, respectively. The corresponding frequency representation using the 3D DFT can be given by

$$G(l,m,n) = \frac{1}{\sqrt{LMN}} \sum_{x=0}^{L-1} \sum_{y=0}^{M-1} \sum_{z=0}^{N-1} g(x,y,z) \omega_L^{lx} \omega_M^{my} \omega_N^{nz} \quad (7)$$

where $\omega_L = e^{-j2\pi/L}$, $\omega_M = e^{-j2\pi/M}$, and $\omega_N = e^{-j2\pi/N}$. Thus, similar to 1D image sets, an orthonormal basis for the image data matrix X_3 can be generated using the basis for the 3D DFT. In particular, the following matrix represents one 3D frequency:

$$F_{\alpha\beta\gamma}(x,y,z) = \frac{1}{\sqrt{LMN}} \omega_L^{\alpha x} \omega_M^{\beta y} \omega_N^{\gamma z} \quad (8)$$

where α, β, and γ denote the desired frequency components in three dimensions with $0 \leq x \leq L-1$, $0 \leq y \leq M-1$, and $0 \leq z \leq N-1$. All $F_{\alpha\beta\gamma}$ matrices can be lexicographically ordered (similar to the ordering of $\mathbf{x}_{lmn}$ vectors in X_3) into their respective column vectors (denoted $\mathbf{f}_{\alpha\beta\gamma}$) so that the corresponding $(L \times M \times N) \times (L \times M \times N)$ "3D" Fourier matrix is given by

$$F_3 = [\mathbf{f}_{000} | \cdots | \mathbf{f}_{\alpha\beta\gamma} | \cdots | \mathbf{f}_{(L-1)(M-1)(N-1)}]. \quad (9)$$

Note that the columns of F_3 give the 3D DFT basis for complex matrices. However, for real images in 3D image sets, the matrix X_3 in (6) will contain all real values and hence, similar to the basis given by columns of H in (4) for 1D image sets, X_3 will have a real basis. To find this real basis, Euler's formula ($e^{-jx} = \cos x - j \sin x$) can be used to rewrite (8) as follows:

$$\begin{aligned} F_{\alpha\beta\gamma}(x,y,z) = \\ \frac{1}{\sqrt{LMN}} [& (c_{\alpha x} c_{\beta y} c_{\gamma z} - c_{\alpha x} s_{\beta y} s_{\gamma z} \\ & - s_{\alpha x} c_{\beta y} s_{\gamma z} - s_{\alpha x} s_{\beta y} c_{\gamma z}) \\ & - j(c_{\alpha x} c_{\beta y} s_{\gamma z} - c_{\alpha x} s_{\beta y} c_{\gamma z} \\ & - s_{\alpha x} c_{\beta y} c_{\gamma z} + s_{\alpha x} s_{\beta y} s_{\gamma z})] \end{aligned} \quad (10)$$

where $c_{\alpha x} = \cos(\frac{2\pi\alpha x}{L})$, $c_{\beta y} = \cos(\frac{2\pi\beta y}{M})$, and $c_{\gamma z} = \cos(\frac{2\pi\gamma z}{N})$, while $s_{\alpha x}$, $s_{\beta y}$, and $s_{\gamma z}$ are the corresponding sine components. Let r denote the number of non-zero α, β, and γ frequencies. Then there will be 2^r different sine-cosine combinations for $F_{\alpha\beta\gamma}$. If these sine-cosine combinations are lexicographically ordered and are scaled by $\sqrt{2^r}$ to give orthonormal columns of $H_{\alpha\beta\gamma}$, then the *real* 3D Fourier matrix of size $(L \times M \times N) \times (L \times M \times N)$ can be given by

$$H_3 = \left[\begin{array}{cccc} \mathbf{f}_{000} & \cdots & H_{\alpha\beta\gamma} & \cdots \end{array} \right] \quad (11)$$

where the first column, $\mathbf{f}_{000}$, of H_3 refers to the DC component corresponding to $r = 0$. Note that if any of the three dimensions, for e.g., L, is even, then only the cosine (real) component of the corresponding maximum "real" frequency, i.e., $\frac{L}{2}$, is considered, otherwise both cosine (real) and sine (imaginary) components of $\frac{L}{2}$ are considered while generating the corresponding orthonormal columns in H_3.

The resulting matrix H_3, which is generated for a given X_3, can be used to extend Chang's algorithm to compute the approximate SVD of X_3. In particular, the row space of X_3 can be projected to the first few columns of H_3 and the SVD of $X_3 H_{3(p)}$ can be used to approximate the SVD of X, where $H_{3(p)}$ denotes the matrix containing the first p columns of H_3.

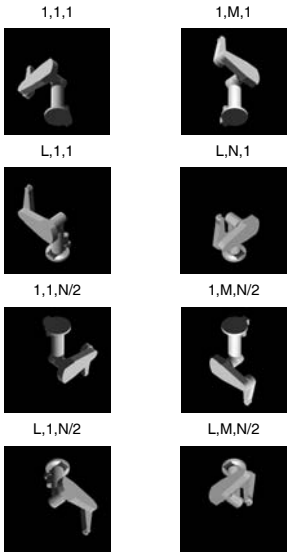

Figure 2. This figure shows a few example synthetic images of a PUMA 560 Robot generated using ray-tracing. The first two letters in the heading indicate the simulated camera location coordinates, while the third letter indicates the image plane rotation index.

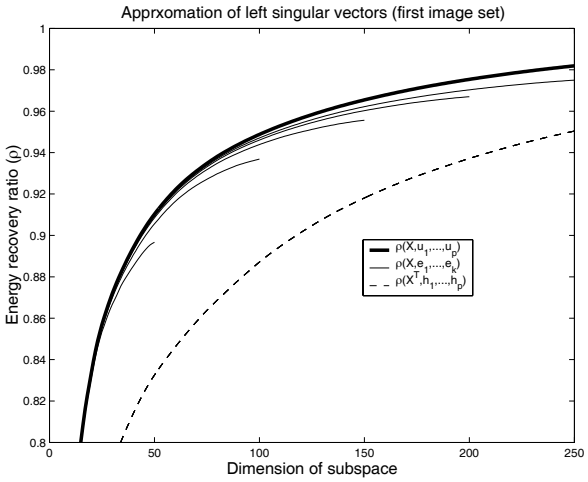

Figure 3. This figure shows the typical relationship between the true left singular vectors, the computed estimates (as a function of k, $1 \leq k \leq p$, for several fixed values of p), and pure sinusoids. The plots shown here are for an image data matrix generated for the robot in Fig. 2 with $L = M = N = 10$.

5 Experimental results

The proposed extension of Chang's algorithm to efficiently compute the SVD of fully general 3D image sets was analyzed using a variety of objects. Two ray-traced image sets of a PUMA 560 robot are used as illustrative examples here. The first image set had the same number of images in all three dimensions with $L = M = N = 10$, while the second image set used $L = M = 10$, $N = 36$. Recall that the simulation setup used in this study considers a range of 90 degrees for both α_l and β_m (refer to Fig. 1), while the γ_n parameter is allowed to span a full 360 degrees. Therefore, the angular separation between consecutive images in the first image set is different, while it is kept the same in the second image set.

Fig. 2 illustrates a few of the images (with $64 \times 64 = 4096$ pixels each) showing different views of the robot taken from simulated camera locations that are at the four extreme corners of the spherical patch above the robot. In particular, images in the last two rows in Fig. 2 are obtained after rotating the corresponding images in the first two rows through 180 degrees. Using the two image sets, the corresponding image data matrices X_3 and the real Fourier matrices H_3 were generated. In particular, X_3 and H_3 for the first image set were of size 4096×1000 and 1000×1000, respectively, while for the second image set, they were of size 4096×3600 and 3600×3600, respectively.

For both image sets, the columns of H_3 were placed in descending order of their ability to recover energy in X_3 using $\rho(X^T, \mathbf{h}_1, \mathbf{h}_2, \cdots)$ in (2). It was observed that these ordered harmonics can be effectively used to improve the computational efficiency of the SVD of X_3. This is illustrated in Fig. 3 and Fig. 4, where ρ is again used to measure the quality of the estimates of the $\mathbf{u}_i$. The solid line shows $\rho(X, \mathbf{u}_1, \cdots, \mathbf{u}_p)$ as a function of p, while the dotted lines show $\rho(X, \mathbf{e}_1, \cdots, \mathbf{e}_k)$ for $k = 1, 2, \cdots, p$ and $p = 50, 100, 150, 200, 250$. It is evident that the $\mathbf{e}_i$ give good estimates of the $\mathbf{u}_i$ even at small values of p indicating an effective extension of Chang's algorithm. Note that the difference between the results for both image sets is minimal considering the difference between the number of images used in the corresponding X_3 matrices.

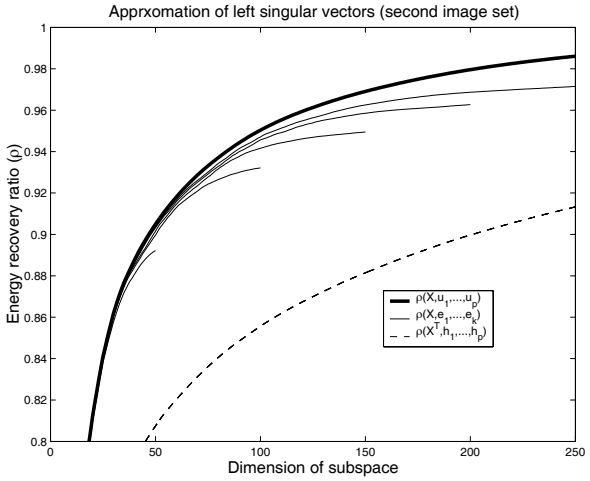

Figure 4. This figure shows the same plots as in Fig. 3 for an image data matrix generated for the robot in Fig. 2 with $L = M = 10$ and $N = 36$.

6 Conclusion and Future Work

This paper has illustrated an extension of Chang's algorithm to compute an estimate of the SVD of 3D image sets, in which the images are characterized by three parameters instead of one. The empirical evidence shows that this extension can be effectively used to improve the offline computational efficiency of calculating the SVD of 3D image sets commonly used in 3D pose estimation.

The work presented here lays a foundation for a computationally efficient SVD algorithm for 3D image sets. However, there are still two issues that need to be resolved. The first issue is related to the optimum ordering of the frequencies of 3D image sets based on their energy recovery ability, while the second issue involves the choice of the number of frequencies, p that should be considered before computing the SVD of $X_3 H_{3(p)}$. For 1D image sets, it was observed that the columns of H in (4) were already ordered in terms of their ability to recover energy in X and $\rho(X^T, \mathbf{h}_1, \cdots, \mathbf{h}_p)$ provided a very conservative lower bound for $\rho(X, \mathbf{e}_1, \cdots, \mathbf{e}_k)$ and hence the two issues were automatically resolved in Chang's algorithm [10]. However, because the columns of H_3 are unordered and because the corresponding $\rho(X_3^T, \mathbf{h}_1, \cdots, \mathbf{h}_p)$ gives a very loose lower bound for $\rho(X_3, \mathbf{e}_1, \cdots, \mathbf{e}_k)$, there is a need to resolve the two aforementioned issues.

7 Acknowledgments

The authors would like to acknowledge Bryce Eldridge for his help with the generation of ray-traced images for this paper.

References

[1] S. K. Nayar, S. A. Nene, and H. Murase. Subspace method for robot vision. *IEEE Trans. Robot. Automat.*, 12(5):750–758, Oct. 1996.

[2] H. Murase and S. K. Nayar. Detection of 3D objects in cluttered scenes using hierarchical eigenspace. *Pattern Recognit. Lett.*, 18(4):375–384, April 1997.

[3] M. H. Yang, D. J. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Trans. PAMI*, 24(1):34–58, Jan. 2002.

[4] S. Shlien. A method for computing the partial singular value decomposition. *IEEE Trans. PAMI*, 4(6):671–676, Nov. 1982.

[5] H. Murakami and V. Kumar. Efficient calculation of primary images from a set of images. *IEEE Trans. PAMI*, 4(5):511–515, Sept. 1982.

[6] R. Haimi-Cohen and A. Cohen. Gradient-type algorithms for partial singular value decomposition. *IEEE Trans. PAMI*, PAMI-9(1):137–142, Jan. 1987.

[7] C. R. Vogel and J. G. Wade. Iterative SVD-based methods for ill-posed problems. *SIAM J. Sci. Comput.*, 15(3):736–754, May 1994.

[8] H. Murase and M. Lindenbaum. Partial eigenvalue decomposition of large images using the spatial temporal adaptive method. *IEEE Trans. Image Processing*, 4(5):620–629, May 1995.

[9] S. Chandrasekaran, B. Manjunath, Y. Wang, J. Winkeler, and H. Zhang. An eigenspace update algorithm for image analysis. *CVGIP: Graphic Models and Image Processing*, 59(5):321–332, Sept. 1997.

[10] C. Y. Chang, A. A. Maciejewski, and V. Balakrishnan. Fast eigenspace decomposition of correlated images. *IEEE Trans. Image Processing*, 9(11):1937–1949, Nov. 2000.