Aerial Pose Detection of 3-D Objects Using Hemispherical Harmonics*

Randy C. Hoover and Anthony A. Maciejewski
Dept. of Electrical and Computer Eng.
Colorado State University
Fort Collins, CO 80523-1373, USA
Email: {hooverr, 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

In this paper, we consider pose estimation of 3-D objects from an aerial perspective using eigendecomposition. We first outline a sampling method to acquire images of the object from an aerial view by sampling the upper unit hemisphere. Using this hemispherical sampling pattern, the computational burden of computing the eigendecomposition can be reduced by using the HemiSpherical Harmonic Transform (HSHT) to "condense" information due to the hemispherical correlation. We then propose a computationally efficient algorithm for approximating the eigendecomposition based on the HSHT analysis.

1. Introduction

Over the last several decades, object recognition and pose detection of three-dimensional (3-D) objects from two-dimensional (2-D) images have become important issues in computer vision and robotics applications. Subspace methods represent one computationally efficient approach for dealing with this class of problems. Subspace methods, also referred to as eigenspace methods, principal component analysis, or the Karhunen-Loeve transformation [1], have been used for a variety of application domains. All of these applications are based on the fact that a set of highly correlated images can be approximately represented by a small set of eigenimages [2, 3]. Once the principal eigenimages of an image data set have been determined, using these eigenimages is very computationally efficient for the on-line classification of 3-D objects. Unfortunately, the off-line calculation for determining the appropriate subspace dimension, as well as the principal eigenimages themselves is computationally expensive.

This drawback has been addressed using several different approaches [4–7]. In [7], Chang *et al.* showed that if the image data set was correlated in one dimension, then the FFT may be used to approximate the required subspace dimension, as well as the principal eigenimages. This result is based on the fact that for a large class of objects, most of the energy in the image data set is contained in the low frequency harmonics.

In this paper the work of [7] is extend to objects correlated in 2-dimensions. In general, if pose detection of 3-D objects from an aerial view is desired, the training image data set needs to contain views of the object captured from an aerial perspective. Acquiring images of the object from an aerial view may be accomplished by sampling the upper unit hemisphere with the object at its center. Once the image data set is constructed in this manner, we take advantage of the hemispherical sampling pattern by using the *HemiSpherical Harmonic Transform* (HSHT) detailed in [8] in place of the FFT to capture the frequency information of this data set. We then propose an algorithm for approximating the subspace dimension, as well as the principal eigenimages using the resulting HSHT analysis.

2. Preliminaries

In this work (as in [7]), a gray-scale image is described by an $h \times v$ array of square pixels with intensity values normalized between 0 and 1. Thus, an image will be represented by a matrix $\mathcal{X} \in [0,1]^{h \times v}$. Because sets of related images are considered in this paper, the *image vector* $\boldsymbol{f}$ of length $m = hv$ is obtained by "row-scanning" an image into a column vector, i.e., $\boldsymbol{f} = \text{vec}(\mathcal{X}^T)$. The *image data matrix* of a set of images $\mathcal{X}_1, \ldots, \mathcal{X}_n$ is an $m \times n$ matrix, denoted X, and defined as $X = [\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n]$, where typically $m > n$ with fixed n. Because we will be sampling images on the unit hemisphere, it should be noted that $\boldsymbol{f} = \boldsymbol{f}(\boldsymbol{\xi}_p)$ where $\boldsymbol{\xi}_p$, $p \in \{0, \ldots, n-1\}$ is the unit vector pointing at the angle of co-latitude $\theta_p \in [0, \pi/2]$ measured down from the z axis, and the angle of longitude $\phi_p \in [0, 2\pi)$, which is the parameterization of the hemisphere in spherical coordinates.

*This work was supported in part by the Missile Defense Agency under contract no. HQ0006-05-C-0035. Approved for Public Release 07-MDA-3043 (28 JAN 08).

The singular value decomposition (SVD) of X is given by

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

where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal, and $\Sigma = [\Sigma_d\ \mathbf{0}]^T \in \mathbb{R}^{m\times n}$ where $\Sigma_d = \text{diag}(\sigma_1, \cdots, \sigma_n)$ 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 $\mathbf{u}_i, i = 1, \ldots, m$, are referred to as the left singular vectors or eigenimages of X, while the columns of V, denoted $\mathbf{v}_i, i = 1, \ldots, n$ are referred to as the right singular vectors of X. In practice, the left singular vectors $\mathbf{u}_i$ are not known or computed exactly, and instead estimates $\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_k$ which form a k-dimensional basis are used. The measure we will use for quantifying the accuracy of these estimates is the "energy recovery ratio" denoted ρ, and is defined as

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

where $||\cdot||_F$ denotes the Frobenius norm. Note that if the $\tilde{\mathbf{u}}_i$ are orthonormal, $\rho \leq 1$ [7].

The principal calculation required with subspace methods is the precomputation of estimates of the left singular vectors $\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_k$ of the $m \times n$ matrix X. This is a very computationally expensive operation when m and n are large. Reducing this computational expense by exploiting correlation between images has been the topic of much previous research. In [7], Chang *et al.* showed that if the image data matrix was correlated in one dimension, then the FFT may be used to approximate the desired subspace dimension k, as well as the principal eigenimages $\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_k$. Our solution to this problem follows the work of Chang; however, because the object is no longer correlated in only one dimension, the direct FFT approach is no longer effective. Furthermore, because we are sampling on the hemisphere, the FFT is multivalued at the north pole and discontinuous at the equator; hence, it is not well posed in the latitude direction. It is still possible however to use the HSHT to approximate the desired subspace dimension k, as well as the principal eigenimages of the image data matrix X. A brief introduction to the HSHT is presented in the next section.

3. Spherical Harmonics

3.1. Introduction

Over the last decade, spherical harmonics have been gaining popularity in the computer vision and computer graphics arena [9–14]. In the context of modern computer vision applications, the information received is in digital format, thus the development of a discrete spherical harmonic transform is needed to process the data. The tessellation of the sphere to form a discrete sampling pattern can be done using several different approaches. In this work, we use the Hierarchical Equal Area isoLatitude Pixelization (HEALPix) [15] sampling pattern. The HEALPix sampling pattern has several advantages over other tessellations, in particular, it does not discriminate by oversampling the polar region [16]. An example of the proposed method for capturing images on the hemisphere using the HEALPix sampling pattern is depicted in Fig. 1.

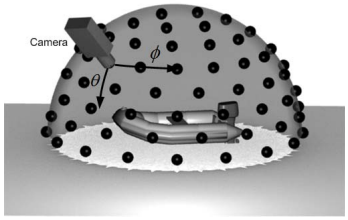

Figure 1. Image acquisition for the training image data matrix X. A sample image is taken at each black dot on the hemisphere.

3.2. Hemispherical Harmonic Transform

A real valued bandlimited function $\boldsymbol{f}(\gamma_p)$ whose domain is the upper hemisphere of $L^2(S^2)$ may be represented by its hemispherical harmonic expansion as

$$f(\boldsymbol{\xi}_p) = \sum_{l=0}^{l_{\max}} \sum_{|m|\leq l} f_l^m H_l^m(\boldsymbol{\xi}_p) \tag{3}$$

where $f(\boldsymbol{\xi}_p) \in [0, 1]$ is a single pixel of the image data vector $\boldsymbol{f}(\boldsymbol{\xi}_p)$. Once again we remind the reader that $\boldsymbol{\xi}_p$, $p \in \{0, \ldots, n-1\}$ is the unit vector pointing at the angle of co-latitude $\theta_p \in [0, \pi/2]$ measured down from the upper pole, and the angle of longitude $\phi_p \in [0, 2\pi)$ is the parameterization of the unit hemisphere in spherical coordinates. In the above equation, it is assumed that the signal power for $l > l_{\max}$ is insignificant, and $l_{\max}$ is chosen such that aliasing does not occur. The expansion coefficients are calculated by

$$f_l^m = \frac{4\pi}{n} \sum_{p=0}^{n-1} f(\boldsymbol{\xi}_p) H_l^m(\boldsymbol{\xi}_p), \tag{4}$$

where $H_l^m(\boldsymbol{\xi}_p)$ is the real-valued hemispherical harmonic defined by

$$H_l^m(\boldsymbol{\xi}_p) = \begin{cases} \sqrt{2}\tilde{\kappa}_{l,m} \cos(m\phi_p) \tilde{P}_l^m(x) & \text{if } m > 0 \\ \sqrt{2}\tilde{\kappa}_{l,m} \sin(|m|\phi_p) \tilde{P}_l^{|m|}(x) & \text{if } m < 0 \\ \tilde{\kappa}_{l,0} \tilde{P}_l^0(x) & \text{if } m = 0 \end{cases} \tag{5}$$

with $\tilde{P}_l^m(x)$ being a shifted associated Legendre polynomial of degree l and order m, $x = 2\cos(\theta_p) - 1$, and

$$\tilde{\kappa}_{l,m} = \sqrt{\frac{2l+1}{2\pi}\frac{(l-|m|)!}{(l+|m|)!}} \quad (6)$$

is a normalization constant [8].

4. Eigendecomposition Algorithm

Our objective is to estimate the desired subspace dimension k, as well as the principal eigenimages $\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_k$ of X such that $\rho(X, \tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_k) > \mu$, where μ is the user specified energy recovery ratio. The first step in computing the desired subspace dimension k, as well as the principal eigenimages, is to construct the image data matrix X. As mentioned in Section 3, the HEALPix sampling pattern is used to capture images of the object in the upper hemispherical region (Fig. 1). This sampling pattern is based on subdividing the sphere using the parameter N_{side}, resulting in $2N_{\text{side}}(3N_{\text{side}}+1)$ images captured from various aerial vantage points on the hemisphere [15]. Setting $l_{\max} = 1.5N_{\text{side}} - 1$, the shifted Legendre polynomials will not alias on the hemisphere for $N_{\text{side}} > 1$ and a power of 2. Furthermore, because the sampling pattern is isolatitudinal, the computation of the Legendre polynomials (which is the most computationally expensive portion in the HSHT algorithm) is minimal.

In this paper, we are using CAD generated ray-traced images, examples of which are shown in Fig. 2 (the CAD models were provided by [17]). Once the image data matrix X has been constructed, we compute the matrix G whose i^{th} row is the HSHT of the i^{th} row of X, denoted from this point forward as HSHT(X). Note, the HSHT(X) may be computed quickly by pre-computing the hemispherical harmonics for different values of N_{side} and storing them for later use. Even though the image data matrix X is hemispherically correlated (i.e., correlated in both θ and ϕ) as opposed to a single parameter, similar to [7], the principal eigenimages $\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_k$ of SVD(G) serve as excellent estimates to those of X. We now propose an algorithm for estimating the required subspace dimension k, as well as the principal eigenimages of X, based on a user specified energy recovery ratio μ.

HSHT Eigendecomposition Algorithm

1. Form the matrix G which is the HSHT(X).
2. Form the matrix H whose columns are the ordered columns of G in descending order according to their norm.
3. Set $q = (1.5N_{\text{side}})^2[1-(1/2)^{N+1}]$, with N=0 initially.
4. Construct the matrix H_q which is the matrix consisting of the first q columns of H.

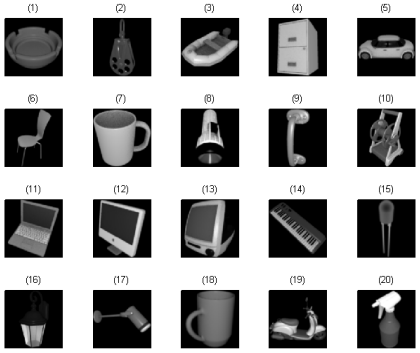

Figure 2. Ray-traced CAD models courtesy of Kator Legaz [17]. Each object is sampled as discussed above at a resolution of 128 × 128. Each of the images are both scale and intensity normalized.

5. Compute SVD(H_q). The key observation here is that H_q contains q columns which is considerably less than the $2N_{\text{side}}(3N_{\text{side}}+1)$ columns of X.
6. If $\rho(X, \tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_q) < \mu$. Let $N = N + 1$ and repeat Steps 3 through 6.
7. Return $\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_k$ such that $\rho(X, \tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_k) \geq \mu$. Note that $k \leq q$.

It should be noted that with the above truncation of $l_{\max}$, only $(1.5N_{\text{side}})^2$ harmonic images are generated, as a result, it is not possible to recover 100% of the energy. While this may be viewed as a drawback in terms of image reconstruction, this is actually an advantage in terms of computational complexity. Furthermore, because subspace methods typically work well with a much smaller subspace dimension, and as a result less recoverable energy, this method has shown significant computational savings with little loss in performance. The actual amount of recoverable energy is dependant on the object; however for all 20 objects in Fig. 2, over 92% of the energy was recoverable at $N_{\text{side}} = 8$.

5. Experimental Results

The above HSHT eigendecomposition algorithm was tested on the objects shown in Fig. 2. The parameter $N_{\text{side}} = 8$ was used, resulting in $2N_{\text{side}}(3N_{\text{side}}+1) = 400$ images per object. The images were then both scale and intensity normalized to create the image data matrix X. Finally, the matrix G was computed condensing the image data set from 400 images to $(1.5N_{\text{side}})^2 = 144$ harmonic images. The true SVD(X) was also computed for a ground truth comparison.

Table 1 shows the mean, min, and max time required to estimate the subspace dimension k, and return the left singular vectors $\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_k$ required to meet the user specified

energy recovery ratio $\mu = 0.85$ and $\mu = 0.9$ for all objects in Fig. 2. As can be seen from the table, the proposed algorithm results in an average speed up of 5.62 for $\mu = 0.85$, and 4.65 for $\mu = 0.90$ as compared to the direct SVD(X). In the table, all times are in seconds computed with MATLAB on a 2.13 GHz Intel dual core. The subspace dimension required for the proposed algorithm to meet the user specified energy recovery ratio $\mu = 0.85$ and $\mu = 0.9$ is shown in Fig. 3. As can be seen in the figure, the estimated subspace dimension is typically not much larger than that computed by the direct SVD(X), specifically at the lower value of μ.

Table 1. Computation time for the proposed algorithm as compared to SVD(X) for all objects in Fig. 2.

	True	Proposed	
	$\mu = 0.85$	$\mu = 0.85$	$\mu = 0.9$
mean	6.69	1.19	1.44
min	6.28	1.05	1.06
max	7.01	1.54	3.30*

* This increased time shows that more than one split was required in step 3 of the HSHT algorithm.

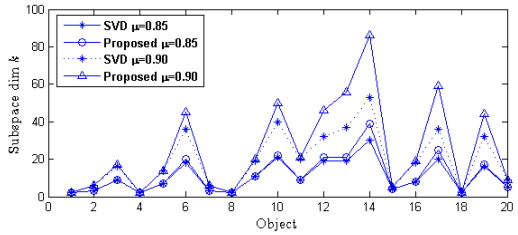

Figure 3. This figure shows the subspace dimension required to meet the user specified energy recovery ratio $\mu = 0.85$ and $\mu = 0.9$.

6. Conclusions and Future Work

We have illustrated a computationally efficient algorithm for estimating the eigendecomposition of hemispherically-correlated images using the discrete hemispherical harmonic transform. The algorithm was then tested on a variety of 3-D objects with images captured from different vantage points around the hemisphere. In addition to significant computational savings as compared to the direct SVD approach, we have shown that the estimated subspace dimension is typically not much larger than the direct SVD approach. Future work will focus on validating the proposed algorithms on true 3-D objects rather than CAD models.

References

[1] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. London, U.K.: Academic, 1990.

[2] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance," *Int. J. Comp. Vis.*, vol. 14, no. 1, pp. 5–24, Jan. 1995.

[3] H. Murakami and V. Kumar, "Efficient calculation of primary images from a set of images," *IEEE Trans. PAMI*, vol. 4, no. 5, pp. 511–515, Sept. 1982.

[4] C. R. Vogel and J. G. Wade, "Iterative SVD-based methods for ill-posed problems," *SIAM J. Sci. Comput.*, vol. 15, no. 3, pp. 736–754, May 1994.

[5] S. Chandrasekaran, B. Manjunath, Y. Wang, J. Winkeler, and H. Zhang, "An eigenspace update algorithm for image analysis," *CVGIP: Graphic Models and Image Proc.*, vol. 59, no. 5, pp. 321–332, Sept. 1997.

[6] K. Saitwal, A. A. Maciejewski, R. G. Roberts, and B. Draper, "Using the low-resolution properties of correlated images to improve the computational efficiency of eigenspace decomposition," *IEEE Trans. Image Proc.*, vol. 15, no. 8, pp. 2376–2387, Aug. 2006.

[7] C. Y. Chang, A. A. Maciejewski, and V. Balakrishnan, "Fast eigenspace decomposition of correlated images," *IEEE Trans. Image Proc.*, vol. 9, no. 11, pp. 1937–1949, Nov. 2000.

[8] P. Gautron, J. Křivánek, S.Pattanaik, and K. Bouatouch, "A novel hemispherical basis for accurate and efficient rendering," in *Rendering Techniques 2004, Eurographics Symposium on Rendering*, 2004, pp. 321–330.

[9] R. Ramamoorthi, "Analytic PCA construction for theoretical analysis of lighting variability in images of a Lambertian object," *IEEE Trans. PAMI*, vol. 24, no. 10, pp. 1322 – 1333, Oct. 2002.

[10] S. Romdhani, J. Ho, T. Vetter, and D. J. Kriegman, "Face recognition using 3-D models: Pose and illumination," *Proc. of the IEEE*, vol. 294, no. 11, pp. 1977 – 1999, Nov. 2006.

[11] D. V. Vranic, "An improvement of rotation invariant 3d-shape based on functions on concentric spheres," in *IEEE Conf. Image Proc.*, Sept 14-17, 2003, pp. 757–760.

[12] Ameesh Makadia, and Kostas Daniilidis, "Rotation recovery from spherical images without correspondences," *IEEE Trans. PAMI*, vol. 28, no. 7, pp. 1170–1175, July 2006.

[13] L. Qing, S. Shan, W. Gao, "Face relighting with radiance environment maps," in *IEEE Conf. Comp. Vis. and Patt. Rec.*, June 18-20, 2003, pp. 158–165.

[14] L. Zhang and D. Samaras, "Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics," *IEEE Trans. PAMI*, vol. 28, no. 3, pp. 351–363, Mar. 2006.

[15] Górski, K. M. Hivon, E. Banday, A. J. Wandelt, B. D. Hansen, F. K. Reinecke, and M. Bartelmann, "HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere," *The Astrophysical Journal*, vol. 622, pp. 759–771, Apr. 2005.

[16] R. C. Hoover, A. A. Maciejewski, and R. G. Roberts, "An analysis of sphere tessellations for pose estimation of 3-D objects using spherically correlated images," in *IEEE SSIAI*, accepted to appear March 24-27, 2008.

[17] K. Legaz, "3-D model database for use with blender (web)," www.katorlegaz.com/3d_models/index.php, 2007.