

HIGH-RESOLUTION NAVIGATION ON NON-DIFFERENTIABLE IMAGE MANIFOLDS

Michael B. Wakin,^r David L. Donoho,^s Hyeokho Choi,^r and Richard G. Baraniuk^r

^r Department of Electrical and Computer Engineering, Rice University

^s Department of Statistics, Stanford University

ABSTRACT

The images generated by varying the underlying articulation parameters of an object (pose, attitude, light source position, and so on) can be viewed as points on a low-dimensional *image parameter articulation manifold* (IPAM) in a high-dimensional ambient space. In this paper, we develop theory and methods for the inverse problem of estimating, from a given image on or near an IPAM, the underlying parameters that produced it. Our approach is centered on the observation that, while typical image manifolds are not differentiable, they have an intrinsic multiscale geometric structure. In fact, each IPAM has a family of approximate tangent spaces, each one good at a certain resolution. Putting this structural aspect to work, we develop a new algorithm for high-accuracy parameter estimation based on a coarse-to-fine Newton iteration through the family of approximate tangent spaces. We test the algorithm in several idealized registration and pose estimation problems.

1. INTRODUCTION

Multiscale methods have proven particularly useful in three specific image processing tasks: *data compression* [1], *noise removal* [2], and *fast registration/template matching* [3–5]. The last task seems very different from the first two. Indeed, registration is pursued by a different research community, and the common theoretical model [6] that explains the success of multiscale methods in the first two tasks apparently has nothing to do with the empirical basis for fast registration methods.

In this paper we develop a preliminary theoretical framework for understanding the effectiveness of multiscale algorithms for a range of image understanding problems – including image registration – based on the structure of families of unregistered images. We study *image articulation families*, in which a collection of images differ one from the other through the action of some parameter controlling location, pose, lighting, and so on. The problem of recovering such a parameter from image data includes image registration/template matching as a special case.

We view the collection of images formed by such an image articulation family as a *manifold* of images embedded in a high- (actually infinite-) dimensional space. Typically, this manifold turns out to be non-differentiable, which implies that a multiscale structure exists, as opposed to a monoscale structure which would occur if the manifold were differentiable.

We consider the problem of recovering, from a single image or movie, the articulation parameters using this multiscale viewpoint. While estimating an articulation parameter appears like a standard problem in nonlinear estimation that can be solved using calculus, due to the non-differentiability of image manifolds we find that calculus only works to a certain accuracy at a certain

scale. Therefore we propose a multiscale estimation process in which the scale changes as the accuracy demands increase. This in some ways mimics now-standard methods of image registration, but gives them a new quantitative justification. A forthcoming article [7] expands upon this exposition.

2. THE MANIFOLD VIEWPOINT

2.1. Imaging parameter articulation manifolds (IPAMs)

Consider a mathematical model of images as functions $I : \mathbf{R}^2 \mapsto \mathbf{R}$. We are interested in families of images formed by varying a parameter $\theta \in \Theta$. For example, θ could be a translation parameter specifying the location of an object in the scene; or an orientation parameter specifying the pose; or an articulation parameter specifying, for a composite object, the relative placement of mobile components. The image formed with parameter θ is a function $f_\theta : \mathbf{R}^2 \mapsto \mathbf{R}$; the corresponding family is the *imaging parameter articulation manifold* (IPAM) $\mathbf{F} = \{f_\theta : \theta \in \Theta\}$. The equation $I = f_\theta$ is our way of saying that the observed image I is a particular member f_θ of the family, with parameter θ . In all cases we take Θ as an open set in d -dimensional Euclidean space, and we assume that the relation $\theta \mapsto f_\theta$ is one-to-one.

The set \mathbf{F} is a collection of functions, and we suppose that all these functions are square-integrable: $\mathbf{F} \subset L^2(\mathbf{R}^2)$. Equipping \mathbf{F} with the L^2 metric, we induce a metric on Θ

$$\mu(\theta^{(0)}, \theta^{(1)}) = \|f_{\theta^{(0)}} - f_{\theta^{(1)}}\|_{L^2}. \quad (2.1)$$

Assuming that $\theta \mapsto f_\theta$ is a continuous mapping for the L^2 metric, (Θ, μ) is a nice metric space. Here are some examples.

Translating disk. Here let f_0 be the indicator function of the unit disk, and let $\Theta = \mathbf{R}^2$ act on the disk according to $f_\theta(x) = f_0(x - \theta)$; see Fig. 1(a). It is easy to see that $\mu(\theta^{(0)}, \theta^{(1)}) = m(\|\theta^{(0)} - \theta^{(1)}\|)$, for a monotone increasing function $m \geq 0$, $m(0) = 0$. In fact, if we let B_x denote the disk centered at $x \in \mathbf{R}^2$, then

$$m(\rho) = \text{Area}(B_{(0,0)} \Delta B_{(\rho,0)})^{1/2},$$

where Δ denotes the symmetric difference (see Fig. 1(b)). Similar models can be set up for translates of the square or another set and for *wedgelets*, which are useful for modeling object boundaries in images [8].

Three-dimensional objects. Our model is not limited to articulations in the plane; we may consider entirely different imaging modalities, such as the photography of a 3-D object. In this case, the object may be subject to rotations ($\Theta = SO(3)$), translations ($\Theta = \mathbf{R}^3$), or a combination of both; the metric simply involves the difference between two rendered images as in (2.1). Figure 5 shows an example involving an icosahedron. Additional articulation parameters, such as camera position or lighting conditions, could also be considered.

This work was supported by NSF, ONR, AFOSR, the Texas Instruments Leadership University Program, and sparked by UCLA/IPAM. Email: {wakin, choi, richb}@rice.edu, donoho@stat.stanford.edu.

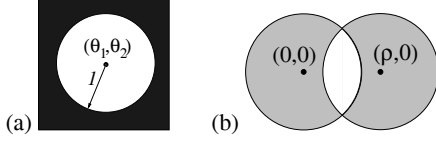


Fig. 1. (a) Parameterization of articulated disk image f_θ . (b) Symmetric difference (shaded) between unit disk and shifted version.

A range of similar models is discussed in [9, 10]; the most elaborate such involve combining some of the above models to create, for example, articulating cartoon faces.

2.2. Non-differentiability

The metric spaces given above all have a non-Lipschitz relation between the metric distance and the Euclidean distance. As one can check by detailed computations [10], we have

$$\mu(\theta^{(0)}, \theta^{(1)}) \geq c \|\theta^{(0)} - \theta^{(1)}\|_2^{1/2} \quad \text{as } \mu \rightarrow 0.$$

The exponent $1/2$ – rather than 1 – implies that the parametrization $\theta \mapsto f_\theta$ is not differentiable. This failure is not something removable by mere reparametrization; no parametrization exists under which there would be a differentiable relationship.

We can view this geometrically. The metric space $M = (\Theta, \mu)$ is isometric to $\mathbf{F} = (\mathbf{F}, \|\cdot\|_{L^2})$. \mathbf{F} is not a smooth manifold; there simply is no system of charts that can make \mathbf{F} even a C^1 manifold. At base, the lack of differentiability of the manifold \mathbf{F} is due to the lack of *spatial* differentiability of these images [10]. In brief, images have edges, and if the locations of edges move as the parameters change then the manifold is not smooth.

2.3. Approximate tangent planes by local PCA

An intrinsic way to think about non-smoothness is to consider approximate tangent planes generated by local principal component analysis (PCA). Suppose we pick an ϵ -neighborhood of some $\theta^{(0)} \in \Theta$; this induces a neighborhood $N_\epsilon(f_{\theta^{(0)}})$ around the point $f_{\theta^{(0)}} \in \mathbf{F}$. We define the ϵ -tangent plane to \mathbf{F} at $f_{\theta^{(0)}}$ as follows. We place a uniform probability measure on θ , inducing a measure ν on the neighborhood $N_\epsilon(f_{\theta^{(0)}})$. We take the first d “principal components” of the “point cloud” defined by random sampling from \mathbf{F} according to ν on the ϵ -neighborhood. The resulting tangent plane $T_{f_{\theta^{(0)}}}^\epsilon(F)$ is simply an approximate least-squares fit to the manifold over the neighborhood $N_\epsilon(f_{\theta^{(0)}})$.

If the manifold were differentiable, then the approximate tangent planes $T_{f_{\theta^{(0)}}}^\epsilon(F)$ would converge to a fixed d -dimensional space as $\epsilon \rightarrow 0$; namely, the plane spanned by the d directional derivatives $\frac{\partial}{\partial \theta_i} f_\theta \big|_{\theta=\theta^{(0)}}$, $i = 1, 2, \dots, d$. However, when these do not exist, the approximate tangent planes do not converge as $\epsilon \rightarrow 0$, but continually “twist off” into other dimensions.

As an example, consider the “translating disk” model, so that the underlying parametrization is 2-D and the tangent planes are 2-D as well. Figure 2(a) shows the approximate tangent plane obtained from this approach, at scale $\epsilon = 1/4$. The tangent plane has a basis consisting of two elements, each of which can be considered an image. Figure 2(b) shows the tangent plane basis images at the finer scale $\epsilon = 1/8$. It is visually evident that the tangent plane bases at these two scales are different; in fact the angle between the two subspaces is approximately 30° . Moreover, since the basis



Fig. 2. Tangent plane basis vectors estimated using local PCA. (a) Scale $\epsilon = 1/4$. (b) Scale $\epsilon = 1/8$.

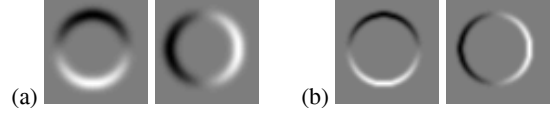


Fig. 3. Tangent plane basis vectors of smoothed manifold \mathbf{F}_s . (a) Scale $s = 1/8$. (b) Scale $s = 1/16$.

elements resemble annuli of shrinking width and growing amplitude, it is apparent for continuous-domain images that as $\epsilon \rightarrow 0$, the tangent plane bases cannot converge in L^2 .

2.4. Approximate tangent planes by regularization

The “twisting off” phenomenon can be understood as the existence of an intrinsic *multiscale structure* to the manifold. Tangent planes, instead of being associated with a location only, as in traditional monoscale analysis, are now associated with a location and a scale.

For a variety of reasons, it is convenient in formalizing this notion to work with a different notion of approximate tangent plane. We first define the family of regularized manifolds as follows. Associated with a given IPAM, we have a family of *regularization operators* Φ_s that act on functions $f \in \mathbf{F}$ to smooth them; the parameter $s > 0$ is a scale parameter. For example, for the translated disk model, we let Φ_s be the operator of convolution with a Gaussian of standard deviation s , $\Phi_s f = \phi_s \star f$. We also define $f_{\theta,s} = \Phi_s f_\theta$. The functions $f_{\theta,s}$ are smooth, and we obtain, for each $s > 0$, a manifold \mathbf{F}_s . The operator family Φ_s should have the property that, as we smooth less, we do less: $\Phi_s f_\theta \rightarrow_{L^2} f_\theta$, $s \rightarrow 0$. It follows that, at least on compact subsets of \mathbf{F} ,

$$\mathbf{F}_s \rightarrow_{L^2} \mathbf{F}, \quad s \rightarrow 0. \quad (2.2)$$

Definition 2.1 The *approximate tangent plane at scale $s > 0$* $T(s, \theta^{(0)}; \mathbf{F})$ is the exact tangent plane of the approximate manifold \mathbf{F}_s : $T_{f_{\theta^{(0)},s}}(\mathbf{F}_s)$.

$T(s, \theta^{(0)})$ is the affine span of the functions $\frac{\partial}{\partial \theta_i} f_{\theta,s} \big|_{\theta=\theta^{(0)}}$, $i = 1, 2, \dots, d$. This notion of approximate tangent plane is different than the more intrinsic local PCA approach but is far more amenable to analysis and computation. In practice, the two notions are similar. Convolution averages an image f_θ with shifted versions of itself; for the translated disk model, these shifted versions simply correspond to nearby points on the manifold. Taking the derivative of this “averaged” manifold is analogous to the local PCA technique; [7] discusses this connection in more depth.

As an example, consider again the “translating disk” model. Figure 3(a) shows the tangent plane obtained from this approach, at scale $s = 1/8$. Figure 3(b) shows the tangent plane at a finer scale, $s = 1/16$. It is again visually evident that the tangent plane bases at the two scales are different, with behavior analogous to the bases shown in Fig. 2.

3. HIGH-RESOLUTION PARAMETER ESTIMATION

With the multiscale viewpoint as background, we now consider the problem of inferring parameters from individual images.

3.1. The problem

We let $\{f_\theta : \theta \in \Theta\}$ be an articulation family. We are given an image I that is known to be of the form $I = f_\theta$ for an unknown $\theta \in \Theta$. We aim to recover θ from I , so conceptually we seek a procedure Q that generates parameter estimates $\hat{\theta} = Q(I)$. There is also a noisy version of the problem, where $I \approx f_\theta$, and we wish to recover an approximation to θ . We put the issue of noise aside for the moment.

A natural approach to the problem is the method of nonlinear least-squares. We pose the objective function $J(\theta) = \|f_\theta - I\|_2^2$, and we seek the minimizing θ : $Q(I) = \operatorname{argmin}_\theta J(\theta)$, supposing that the minimum is uniquely defined.

3.2. Inspiration

Standard nonlinear parameter estimation [11] tells us that, if J is differentiable, then we can use calculus to refine an initial guess $\theta^{(0)}$ at the unknown parameter; we take a Newton step from that guess in the direction of the minimum. Moreover, if the initial guess is reasonably good, then the step from $\theta^{(0)}$ to $\theta^{(1)}$ will place us dramatically closer to the minimum, squaring the error. Continuing with $\theta^{(2)}$, $\theta^{(3)}$, etc., we obtain superlinear convergence.

In the case of a *differentiable* manifold, iteration $k + 1$ of this algorithm would proceed as follows:

1. Compute/estimate the local tangent vectors $v_i^{(k)} = \frac{\partial}{\partial \theta_i} f_\theta |_{\theta=\theta^{(k)}}$, $i = 1, 2, \dots, d$.
2. Find the orthogonal projection of the estimation error onto this tangent space, and expand the projection in terms of the tangent vectors $e^{(k)} = \operatorname{Proj}((I - f_{\theta^{(k)}}) \rightarrow T_{f_{\theta^{(k)}}}(F)) = \sum_{i=1}^d \alpha_i v_i^{(k)}$.
3. Use the expansion coefficients to update the estimate $\theta_i^{(k+1)} \leftarrow \theta_i^{(k)} + \alpha_i$, $i = 1, 2, \dots, d$.

In our setting, however, the tangent vectors $v_i^{(k)}$ do not exist, making it impossible to implement such an algorithm. We turn again to the regularization process in order to remedy this situation.

3.3. Multiscale Newton algorithm

As discussed above, the problem of differentiability can be alleviated by regularizing the images f_θ . Thus, navigation is possible on any of the regularized manifolds \mathbf{F}_s using Newton's method as described above. This fact, in conjunction with the convergence property (2.2), suggests a multiscale technique for parameter estimation.

The idea is to select a sequence of scales $s_0 > s_1 > \dots > s_K$, and to start with an initial guess $\theta^{(0)}$. At each scale we take a Newton step on the corresponding smoothed manifold. In particular, iteration $k + 1$ of the algorithm would proceed as follows:

1. Compute the local tangent vectors $v_{i,s_k}^{(k)}$ on the smoothed manifold \mathbf{F}_{s_k} at the point $f_{\theta^{(k)},s_k}$.
2. Project the estimation error $I_{s_k} - f_{\theta^{(k)},s_k}$ (relative to the *regularized* image $I_{s_k} = \phi_{s_k} \star I$) onto the tangent space $T(s_k, \theta^{(k)})$, and expand this in terms of the tangent vectors.

Table 1. Estimation errors of multiscale Newton iterations, translating disk, no noise.

s	θ_1 error	θ_2 error	image MSE
Initial	-1.53e-01	1.92e-01	9.75e-02
1/2	-2.98e-02	5.59e-02	3.05e-02
1/4	-4.50e-04	1.39e-03	1.95e-04
1/16	-1.08e-06	8.62e-07	8.29e-10
1/256	1.53e-08	1.55e-07	1.01e-10

Table 2. Estimation errors of multiscale Newton iterations, translating disk, with noise. MSE between noisy image and true disk = 4.0237.

s	θ_1 error	θ_2 error	image MSE
Initial	-1.53e-01	1.92e-01	4.1262
1/2	-1.92e-02	4.30e-02	4.0427
1/4	7.45e-04	1.66e-03	4.0241
1/16	-1.21e-03	4.63e-03	4.0255
1/256	9.08e-04	1.32e-03	4.0239

3. Use the expansion coefficients to update the estimate and obtain $\theta^{(k+1)}$.

Under certain conditions on the accuracy of the initial guess and the sequence $\{s_k\}$ it can be shown that this algorithm provides estimation accuracy $\|\theta - \theta^{(k)}\| < cs_k^2$. Ideally, we would be able to square the scale between successive iterations, $s_{k+1} = s_k^2$; a more detailed discussion is included in [7].

4. EXAMPLES

4.1. Translating disk

As a basic exercise of the proposed algorithm, we attempt to estimate the articulation parameters for a translated disk. The process is illustrated in Fig. 4. The observed image I is shown at the top; the leftmost image on the second row is the initial guess $f_{\theta^{(0)}}$. For this experiment, we create 256×256 images with "subpixel" accuracy (each pixel is assigned a value based on the proportion of its support that overlaps the disk).

We ran the multiscale estimation algorithm using the sequence of stepsizes $s = 1/2, 1/4, 1/16, 1/256$. Fig. 4 shows the basic computations of each iteration. Note the geometric significance of the smoothed difference images $I_s - f_{\theta^{(k)},s}$; at each scale this image is projected onto the tangent plane basis vectors. Table 1 gives the estimation errors at each iteration, both for the articulation parameters θ and the mean square error (MSE) of the estimated image. Using this sequence of scales, we observe rapid convergence to the correct articulation parameters with accuracy far better than the width of a pixel, $1/256 \approx 3.91e - 03$.

We now run a similar experiment for the case where the observation $I = f_\theta + n$, where n consists of additive white Gaussian noise of variance 4. Using the same sequence of smoothing filter sizes, the results are given in Table 2. Note that the estimated articulation parameters are approximately the best possible, since the resulting MSE is approximately equal to the noise energy.

4.2. 3-D object motion

We now simulate the photography of a 3d icosahedron. Our image model includes a directional light source (with location and intensity parameters assumed known) and a known camera position. In

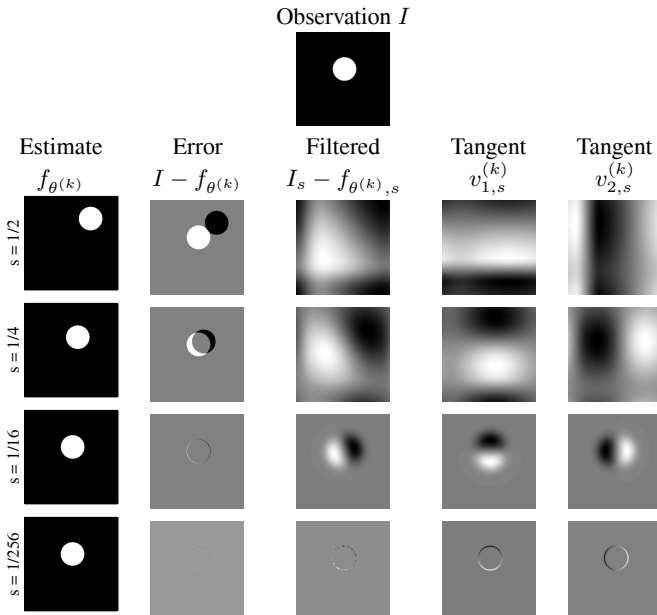


Fig. 4. Multiscale estimation of translation parameters for observed disk image. Each row corresponds to the smoothing and tangent basis vectors for one iteration.

this experiment, the parameter space Θ is 6-D; the object articulations are 3 rotational coordinates and 3 translation parameters. We consider color images, treating each image as an element of $\mathbf{R}^{256 \times 256 \times 3}$. Figure 5 shows the successful estimation of the articulation parameters for a noisy image. For this example, we must use a slightly less ambitious sequence of smoothing filters.

5. DISCUSSION AND CONCLUSIONS

Our multiscale framework for estimation with IPAMs shares common features with a number of practical image registration algorithms; space considerations permit discussion of only a few here. Irani and Peleg [3] have developed a popular multiscale algorithm for registering an image $I(x)$ with a translated and rotated version for the purposes of super-resolution. They employ a multiscale pyramid to speed up the algorithm and to improve accuracy, but a clear connection is not made with the non-differentiability of the corresponding IPAM. While Irani and Peleg compute the tangent basis images with respect to the x_1 and x_2 axes of the image, Keller and Averbach [4] compute them with respect to changes in each of the registration parameters. They also use a multiscale pyramid and conduct a thorough convergence analysis. Belhumeur [5] develops a tangent-based algorithm that estimates not only the pose of a 3-D object, but also its illumination parameters.

In addition to the convergence analysis mentioned in Sec. 3.3, a number of issues remain open. For instance, with noisy images the multiscale tangent projections will reach a point of diminishing returns where finer scales will not benefit; we must develop a stopping criterion for such cases. Additional issues revolve around efficient implementation. We believe that a sampling of the tangent planes needed for the projections can be precomputed and stored using the multiscale representation of [12]. Moreover, since many of the computations are local (as evidenced by the support of the tangent basis images in Figs. 2 and 3), we expect that the image projection computations can be implemented in the wavelet

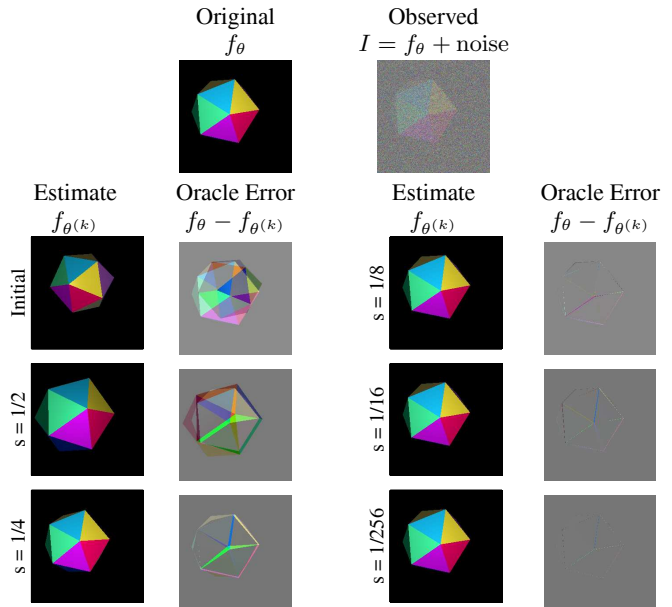


Fig. 5. Estimation of articulation parameters for 3-D icosahedron.

domain. This would also lead to a fast method for obtaining the initial guess $\theta^{(0)}$ with the required accuracy. We will discuss these points and more in greater depth in a forthcoming article [7].

6. REFERENCES

- [1] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2435–2476, October 1998.
- [2] D. L. Donoho, "Denosing by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
- [3] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 3, pp. 231–239, May 1991.
- [4] Y. Keller and A. Averbach, "Fast motion estimation using bidirectional gradient methods," *IEEE Trans. Image Processing*, vol. 13, no. 8, pp. 1042–1054, August 2004.
- [5] P. N. Belhumeur and G. D. Hager, "Tracking in 3D: Image variability decomposition for recovering object pose and illumination," *Pattern Analysis and Applications*, vol. 2, pp. 82–91, 1999.
- [6] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Processing*, vol. 9, no. 9, pp. 1532–1546, September 2000.
- [7] M. B. Wakin, R. G. Baraniuk, H. Choi, and D. L. Donoho, "High-resolution navigation on non-differentiable image manifolds," 2004, preprint.
- [8] D. L. Donoho, "Wedgelets: Nearly-minimax estimation of edges," *Annals of Stat.*, vol. 27, pp. 859–897, 1999.
- [9] C. Grimes, *New methods in nonlinear dimensionality reduction*, Ph.D. thesis, Department of Statistics, Stanford University, 2003.
- [10] D. L. Donoho and C. Grimes, "Image manifolds isometric to euclidean space," *J. Math. Imaging and Computer Vision*, 2003, To appear.
- [11] D. M. Bates and D. G. Watts, *Nonlinear Regression Analysis and Its Applications*, John Wiley and Sons, New York, 1988.
- [12] I. Ur Rahman, I. Drori, V. C. Stodden, D. L. Donoho, and P. Schroeder, "Multiscale representations for manifold-valued data," 2004, preprint.