DEFORMATION-BASED NONLINEAR DIMENSION REDUCTION: APPLICATIONS TO NUCLEAR MORPHOMETRY

Gustavo K. Rohde[1] , *Wei Wang*[1], *Tao Peng*[1] , *Robert F. Murphy*[1,2,3]

[1]Center for Bioimage Informatics, Department of Biomedical Engineering
[2]Department of Biological Sciences
[3]Department of Machine Learning
Carnegie Mellon University, Pittsburgh, PA. 15213

ABSTRACT

We describe a new approach for elucidating the nonlinear degrees of freedom in a distribution of shapes depicted in digital images. By combining a deformation-based method for measuring distances between two shape configurations together with multidimensional scaling, a method for determining the number of degrees of freedom in a shape distribution is described. In addition, a method for visualizing the most representative modes of variation (underlying shape parameterization) in a nuclei shape distribution is also presented. The novel approach takes into account the nonlinear nature of shape manifolds and is related to the ISOMAP algorithm. We apply the method to the task of analyzing the shape distribution of HeLa cell nuclei and conclude that approximately three parameters are responsible for their shape variation. Excluding differences in size, translation, and orientation, these are: elongation, bending (concavity), and shifts in mass distribution. In addition, results show that, contrary to common intuition, the most likely nuclear shape configuration is not symmetric.

Index Terms— Nuclear shape analysis, nonlinear, dimension reduction, image registration.

1. INTRODUCTION

Under the rubric of Computational Anatomy [1], image-based studies of biological forms and structures have been vastly applied to characterize representative organ anatomies across states (normal vs. diseased), ages, populations, etc., from high resolution medical image data. Computational algorithms for extracting quantitative morphological information have become popular since these can often be completely automated, and have become capable of providing detailed, high resolution mappings of, and relationships between, different organs and biological structures.

Previous works in computational anatomy have been almost exclusively applied to characterizing morphology of human organs at scales accessible to standard medical imaging modalities such as computed tomography (CT) and magnetic resonance imaging (MR) with the brain and heart receiving most of the focus. Guimond [2] et al propose an iterative, Euclidean averaging-based approach for estimating a brain template or atlas from a collection of images, while Joshi et al [3] and Avants et al [4], following the framework put forth by Grenander and Miller [1], propose atlas estimation algorithms based on the minimization of nonlinear, geodesic distances, on the manifold of diffeomorphisms relating anatomical instances depicted

in images. Related developments include the works of Ashburner et al [5] who use multivariate analysis of covariance, and Rueckert et al [6] who use principal component analysis (PCA), of deformation fields that warp two images into alignment so as to elucidate principal modes of deformation in a population of brain images. Cootes et al. [7] also promote the use of PCA to discern modes of variation from point landmark extracted from image data. Finally, due to the fact that shape spaces are rarely linear, Vaillant et al [8] propose a method for performing nonlinear statistical analysis on the space of diffeomorphisms through PCA on initial momentum equations from geodesic flows. An alternative approach is provided by Fletcher et al. [9] who propose a method for principal geodesic analysis for the study of nonlinear statistics of shape. We note that the above is a sample of representative works in image-based morphometry for biology and medicine. A more complete overview of previous works is provided by Miller [10].

Here we propose a novel method for recovering the underlying parameterization of a shape distribution by combining the large deformation diffeomorphic metric mapping (LDDMM) framework described by Grenander and Miller [1] and the multidimensional scaling (MDS) technique. The significant modes of variation in a population of shapes are determined by first estimating the geodesic distances between morphological exemplars through a greedy algorithm and then applying MDS to the distance matrix obtained. The number of degrees of freedom in the distribution is estimated from the MDS reconstruction error, while the actual modes themselves are displayed as point distributions on a cartesian grid.

We apply the methodology to characterize normal nuclear shape variation in HeLa cells. The precise characterization of nuclear shape distributions is an important, unsolved problem, with numerous applications to medicine and biology. In a broad sense, nuclei can be viewed as the "brain" of cells and nuclear morphology, and associated changes, have been studied in conjunction with cellular movements, numerous pathologies [11, 12], as well as gene expression and protein synthesis [13] to name a few possible applications of our work. We conclude that approximately three modes are required to accurately describe the variation of nuclear shape in the cell population we study: elongation, bending, and shifts in mass distribution. In addition, we compute an approximate mean nuclear shape and show that, contrary to common intuition, it is not symmetric.

The remaining of this paper is organized as follows. In the next section we briefly describe the imaging and preprocessing set up, as well as the greedy LDMM distance measurement, and MDS analysis. Results using a population of HeLa cell nuclei are then shown, followed by a brief summary and conclusion section.

---

2. METHODS

2.1. Acquisition, pre-processing and initialization

In our experiments we have used previously acquired images of HeLa cell nuclei (a total of $N = 82$ images two dimensional images), obtained as described by Boland and Murphy [14]. As our concern in this work is in characterizing the overall shape of cell nuclei, we use a binary version of each image, obtained via the method described in [15], in subsequent analysis. As the study of shapes is normally regarded as understanding variation in form modulo variations in overall orientation, position, and size, given a set of binary images I_k, $k = 1, \cdots, N$, each containing one nucleus, we initialize each image by finding the spatial transformations that minimize the following functional:

$$\sum_{m=1}^{N-1} \sum_{n=m+1}^{N} |I_m(A_m x + r_m) - I_n(A_n x + r_n)|^2 \qquad (1)$$

where A_m are affine transformations containing rotation and isotropic scaling, and r_m are translation vectors. The minimization of (1) is computationally intensive and therefore we resort to the following approximation. Each binary image is first scaled so that its foreground has the same total area. The translation vectors are computed simply by translating the center of mass of the object to the center of the field of view of each image. All images are then rotated to have the same orientation using a principal axis transform. Finally, each image is then "flipped" left to right, and up and down, simply by reversing coordinates, until functional (1) is minimized.

2.2. Large deformation diffeomorphic metric mapping

Following the approach put forth by Grenander and Miller [1] we aim to understand shape distribution-related quantities by analyzing the spatial deformations that map one shape (image) to another. More specifically, we study the set of forms generated by the group of diffeomorphisms (smooth invertible mappings) g acting on a template image $I(g(x))$, thus generating a morphological "orbit" or shape manifold. The diffeomorphism mapping the coordinate sets Ω of two images I_0 and I_1 is computed as the end point (at time T) of a flow associated to a smooth time- dependent vector field v:

$$\frac{\partial g(x; t)}{\partial t} = v\left(g(x; t); t\right)$$

with $g(x, 0) = x$, and v chosen so as to minimize the following geodesic length definition

$$\int_0^T \|Lv(x; t)\| dt \qquad (2)$$

subject to $I_0(g(x; T)) = I_1(x)$. In (2) L denotes a linear differential operator (in this work $L = \alpha \nabla^2 + \gamma I$, with I denoting the identity matrix in two dimensions) and $\| \cdot \|$ the standard L_2 norm for vector fields on Ω. The end point of the differential equation can be guaranteed to be diffeomorphism by ensuring a sufficient amount of smoothnes on the vector fields v [16]. The infimum (over all possible v) of equation (2) defines a metric (a geodesic distance) on the group of diffeomorphisms generating the morphological orbit [1] (shape manifold).

The problem stated above is solved using the minimum energy template estimation approach:
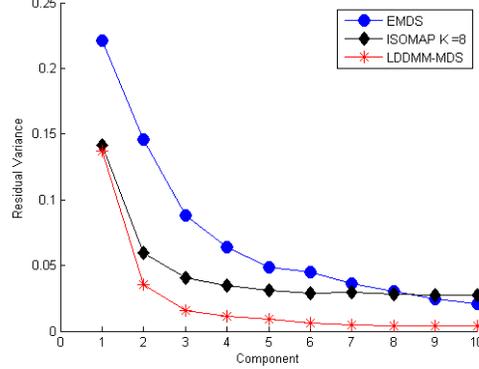
Fig. 1. Residual variance of distance reconstruction as a function of the number of dimension (components) in the reconstruction for LDDMM-MDS, ISOMAP, and EMDS).

$$\int_0^T \|Lv(x; t)\|^2 dt + \int_\Omega |I_0(g(x, T)) - I_1(x)|^2 dx. \qquad (3)$$

The Euler-Lagrange equations for the template estimation problem above have been derived [17]. In this work, however, we use the computationally more efficient locally-in-time optimal "greedy" algorithm described in [1]. Briefly, the final transformation is computed by assuming piecewise constant velocity fields, quantized in time increments $\triangle t$, $t_k = k \triangle t$, $k = 0, \cdots, K$. The locally optimal velocity fields satisfy [1]

$$L^* L v_t + b_t = 0, \qquad (4)$$

with $b_t(x) = -(I_0(g(x; t) - I_1(x)) \nabla I_0(g(x; t))$ and L^* denoting the adjoint of operator L. Following Joshi et al. [3] the deformation field is updated via $g(x; k + 1) = g(x + \epsilon v(x, k), k)$ while the inverse of $L^* L$ was computed with the aid of the Fast Fourier Transform (FFT) as described in [17]. We also use the symmetric, inverse consistent version of the problem stated in equation (3), described in [3, 4]. Briefly, instead of matching image I_0 to I_1 via update equations described above, both images are warped simultaneously to each other, via similar update equations (see [3, 4] for more details).

2.3. Multidimensional scaling

Given a set of such multidimensional points (morphological exemplars), and their pairwise distances computed using the LDDMM framework discussed above, multidimensional scaling (MDS) can be use to finding a low dimensional "Euclidean" embedding that preserves interpoint distances[18], yielding an isometric feature map (see also [19]). Let $D_{m,n} = d^2(I_m, I_n)$, with $d(I_m, I_n)$ representing the distance defined by the infimum of the path integral (equation (2)) generating the spatial transformation $g(x, T)$ warping images I_m, I_n into alignment. The goal in MDS is to find a set of coordinates w_k, $k = 1, \cdots, N$ in a low dimensional Euclidean space that best preserves the notion of distance imparted by the LDDMM framework. This task can be achieved by choosing the top eigenvalues and corresponding eigenvectors of the matrix $G = -0.5(\text{Id} - uu^T)D(\text{Id} - uu^T)$, with $u^T = 1/\sqrt{N}(1, \cdots, 1)$, and Id representing the identity matrix. Let $\lambda_1, \cdots, \lambda_N$ represent the eigenvalues of
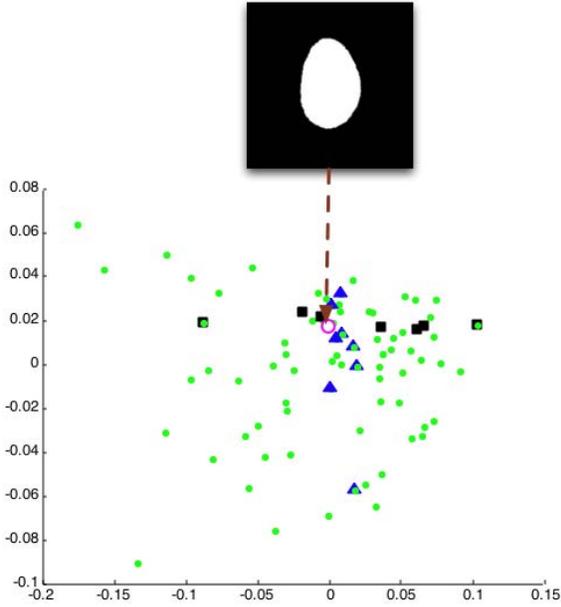
Fig. 2. Two dimensional representation of shape manifold computed using LDDMM-MDS. Each circle represents one image. The points marked with squares and triangles represent two different deformation modes displayed in Figure 3. An approximate average image (see text for definition) is also shown.

G, arranged in decreasing order of magnitude, and with corresponding eigenvectors $g_1, \cdots, g_N$. The ith component of vector w_k is given by $\sqrt{\lambda_i} g_i^k$.

3. RESULTS

3.1. Dimension estimation

Given a set of M-dimensional vectors w_k (constructed from the first M dimensions of each w_k), an approximation $\tilde{D} = \|w_m - w_n\|^2$ of the LDDMM-based distance matrix D described above can be computed. As done in [19], the intrinsic dimensionality of the data (number of free articulation parameters) is estimated by looking for the "elbows" at which the residual variance (defined to be $1 - R^2(\tilde{D}, D)$, with R denoting the correlation coefficient between the entries of both matrices) ceases to decrease significantly with added dimensions (increase in the value of M) in the reconstruction of $\tilde{D}$. Figure 1 displays the residual variance as a function of the number of components (dimensions) used in the distance matrix reconstruction $\tilde{D}$. For comparison purposes, the residual variances for Euclidean MDS (EMDS), MDS applied on the Euclidean pairwise distances between the binary images, as well as the ISOMAP algorithm [19] (the number of neighbors in the graph construction phase in this case was $K = 8$) are included. In all cases, the residual variance decreases as the number of dimension increases. Note that the reconstruction residual variance is worst for EMDS (analogous to PCA in this case [18]). EMDS also estimates a much larger intrinsic dimensionality for the data. Finally, although LDDMM-MDS and ISOMAP seem to estimate the same intrinsic dimensionality for the data (three pa-
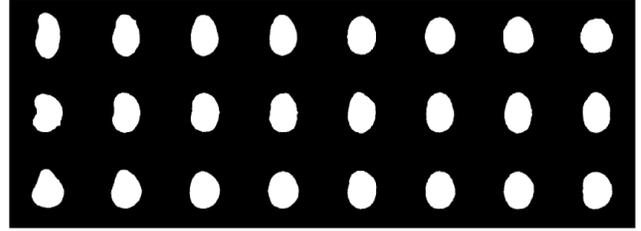
Fig. 3. First three principal modes of deformation (one per row) computed through the LDDMM-MDS framework. In the first mode, each image corresponds to one square shown in Figure 2, in order, from left to right. In the second mode, each image corresponds to one triangle, organized so that the left most image corresponds to the first triangle from the bottom, the second image to the second triangle from the bottom and so on.

rameters), LDDMM-MDS reconstructions contain the least residual variance.

3.2. Deformation modes

A two dimensional projection (the first two components for each vector w_k, $k = 1, \cdots, N$, and N being the number of images) of the (nonlinear) degrees of freedom responsible for the variations in shape is shown in Figure 2. Each small circle corresponds to one image. The images corresponding to the dark squares (variation in the first dimension) are shown in the top row of Figure 3 while the images corresponding to the points marked with triangles (variation along the second dimension) are shown in the second row of Figure 3. The variations along the third dimension of vectors w_k are shown in Figure 3, but omitted from Figure 2 for brevity. See figure captions for more details. The first three modes of deformation clearly correspond to, from most to least significant: elongation, bending, and shifts (asymmetry) in mass distribution.

3.3. Approximate mean

A Frechet mean shape can also be approximated from the LDDMM-based distance matrix D as

$$\bar{I} = \arg\min_{I_k} \frac{1}{N} \sum_{i=1, i \neq k}^{N} D_{k,i}^2. \tag{5}$$

The result is displayed in Figure 2.

4. SUMMARY AND CONCLUSIONS

We have proposed a nonlinear method for characterizing the underlying parameterization in a distribution of shapes. The algorithm proposed relies on a combination of the LDDMM and MDS methods, and is related to the ISOMAP algorithm [19]. As such, it relies on similar assumptions to the ISOMAP algorithm. Most notably, it is assumed that the set of articulation parameters (elongation, bending, etc.) is convex [19]. This assumption could be overly restrictive for some interesting image articulation experiments [20]. Our algorithm, however, does not rely on a nearest neighbor map for computing the pairwise distances, but rather on distances induced on the space of diffeomorphisms connecting two morphological structures.

As such, the method is expected to be more precise, and less dependent on the amount of data available. Experimental results indicate that, even though they both estimate the same number of independent degrees of freedom in this case, LDDMM-MDS produces a lower residual variance reconstruction than ISOMAP.

We also note that the greedy algorithm used in computing geodesic distances is not optimal in the sense that it may produce erroneous (longer) distances. More precise geodesic distances may be computed using the Euler-Lagrange equations associated with the template estimation problem stated in equation (3), albeit at an increase in computational complexity.

The proposed methodology was applied to the analysis of nuclei in HeLa cell microscopic images. With exception of our own prior work [21], we believe the work described here is the first to investigate the application of similar methods to nuclear shape analysis. Registration methods have been recently applied to register images of nuclei [22, 23], but not as an approach to characterize nuclear shape distributions. In summary, the nonlinear analysis of HeLa cell nuclei was able to conclude that, even though each input image contained 196×196 pixels, approximately three parameters are responsible for the variations observed in the dataset: elongation, bending, and shifts in mass distribution. In addition, it was shown that the approximate mean shape is not strictly symmetric (even though the set of images contained many nearly symmetric shapes). This is partly due to the fact the initialization procedure removes variations due to position, size, and orientation (including coordinate inversions). A similar conclusion can be reached by applying a mean template estimation algorithm, such as the one described by Joshi et al [3], to the same dataset [21]. We expect the methodology presented here to play an useful and important role in characterizing cell and nuclear shape for the purposes of teaching, understanding the effects of different drugs and transfection agents, studying cell division and apoptosis, modeling cell behavior, as well as in detection and characterization of pathology, amongst other applications.

5. REFERENCES

[1] U. Grenander and M. I. Miller, "Computational anatomy: an emerging discipline," *Quart. Appl. Math.*, vol. 56, no. 4, pp. 617–694, 1998.

[2] A. Guimond, J. Meunier, and J.-P. Thirion, "Average brain models: a convergence study," *Comput. Vis. Image Underst.*, vol. 77, no. 2, pp. 192–210, 2000.

[3] S. Joshi, B. Davis, M. Jomier, and G. Gerig, "Unbiased diffeomorphic atlas construction for computational anatomy," *Neuroimage*, vol. 23, pp. S151–S160, 2004.

[4] B. Avants and J. C. Gee, "Geodesic estimation for large deformation anatomical shape averaging and interpolation," *Neuroimage*, vol. 23, no. S139-S150, 2004.

[5] J. Ashburner, C. Hutton, R. Frackowiak, I. Johnsrude, C. Price, and K. Friston, "Identifying global anatomical differences: Deformation-based morphometry," *Human Brain Mapping*, vol. 6, pp. 348–357, 1998.

[6] D. Rueckert, A. F. Frangi, and J. A. Schnabel, "Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 1014–1025, 2003.

[7] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models: their training and application," *Comput. Vis. Image Underst.*, vol. 61, no. 1, pp. 38–59, 1995.

[8] M. Vaillant, M.I. Miller, L. Younes, and A. Trouve, "Statistics on diffeomorphisms via tangent space representations," *Neuroimage*, vol. 23, pp. S161–S169, 2004.

[9] P. T. Fletcher, C. L. Lu, S. A. Pizer, and S. Joshi, "Principal geodesic analysis for the study of nonlinear statistics of shape," *IEEE Trans. Med. Imag.*, vol. 23, pp. 995–1005, 2004.

[10] M. I. Miller, "Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms," *Neuroimage*, vol. 23, pp. S19–S33, 2004.

[11] D. Zink, A. H. Fischer, and J. A. Nickerson, "Nuclear structure in cancer cells," *Nat. Rev. Cancer*, vol. 4, no. 677-687, 2004.

[12] K. N. Dahl, P. Scaffidi, M. I. Islam, A. G. Yodh, K. L Wilson, and T. Misteli, "Distinct structural and mechanical properties of the nuclear lamina in Hutchinson-Gilford progeria syndrom," *Proc. Natl. Acad. Sci.*, vol. 103, pp. 10271–10276, 2006.

[13] C.H. Thomas, J. H. Collier, C. S. Sfeir, and K. E. Healy, "Engineering gene expression and protein synthesis by modulation of nuclear shape.," *Proc. Natl. Acad. Sci.*, vol. 99, pp. 1972–1977, 2002.

[14] M.V. Boland and R.F. Murphy, "A neural network classifier capable of recongnizing all major subcellular structures in fluorescence microscope images of hela cells," *Bioinformatics*, vol. 17, pp. 1213–1223, 2001.

[15] N Otsu, "Threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man. Cybern*, vol. 9, pp. 62–66, 1979.

[16] P. Dupuis, U. Grenander, and M. Miller, "Variational problems on flows of diffeomorphisms for image matching," *Quart. Appl. Math.*, vol. 56, pp. 587–600, 1998.

[17] M.F. Beg, M.I. Miller, A. Trouve, and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," *Intern. J. Comp. Vis.*, vol. 61, no. 2, pp. 139–157, 2005.

[18] L.K. Saul, K. Q. Weinberger, J. H. Ham, F. Sha, and D.D. Lee, *Semisupervised Learning*, chapter Spectral methods for dimensionality reduction, MIT Press, 2006.

[19] J.B. Tenembaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, 2000.

[20] D. L. Donoho and C. Grimes, "When does ISOMAP recover the natural parameterization of families of articulated images?," Tech. Rep. 2002-27, Stanford, 2002.

[21] G.K. Rohde, A. J. S. Ribeiro, K. N. Dahl, and R. F. Murphy, "Deformation-based nuclear morphometry: capturing nuclear shape variation in HeLa cells.," *Cytometry A*, in press, 2008.

[22] S. Yang, S. Kohler, K. Teller, T. Cremer, P. Le Baccon, E. Heard, R. Eils, and K. Rohr, "Non-rigid registration of 3d multi-channel microscopy images of cell nuclei.," in *LNCS*, 2006, vol. 4190, pp. 907–914.

[23] J. Mattes, J. Nawroth, P. Boukamp, R. Eils, and K. M. Greulich-Bode, "Analyzing motion and deformation of the cell nucleus for studying co-localizations of nuclear structures," in *Int. Symp. Biomed. Imaging*, 2006, pp. 1044–1047.