The Normal Mode Analysis Shape Detection Method for Automated Shape Determination of Lung Nodules

Joseph N. Stember

Published online: 16 September 2014
© Society for Imaging Informatics in Medicine 2014

Abstract Surface morphology and shape in general are important predictors for the behavior of solid-type lung nodules detected on CT. More broadly, shape analysis is useful in many areas of computer-aided diagnosis and essentially all scientific and engineering disciplines. Automated methods for shape detection have all previously, to the author's knowledge, relied on some sort of geometric measure. I introduce Normal Mode Analysis Shape Detection (NMA-SD), an approach that measures shape indirectly via the motion it would undergo if one imagined the shape to be a pseudomolecule. NMA-SD allows users to visualize internal movements in the imaging object and thereby develop an intuition for which motions are important, and which geometric features give rise to them. This can guide the identification of appropriate classification features to distinguish among classes of interest. I employ normal mode analysis (NMA) to animate pseudomolecules representing simulated lung nodules. Doing so, I am able to assign a testing set of nodules into the classes circular, elliptical, and irregular with roughly 97 % accuracy. This represents a proof-of-principle that one can obtain shape information by treating voxels as pseudoatoms in a pseudomolecule, and analyzing the pseudomolecule's predicted motion.

Keywords Automated measurement · Biomedical image analysis · Classification · Computerized method · Digital image processing

Introduction

Shape characterization and classification is a very broad topic that is of great interest across a wide range of scientific and engineering disciplines. Notably, it has extensive applications to medical imaging. In particular, there is much interest in the role of shape in forecasting the likelihood that small solid-type pulmonary nodules detected on CT will become malignant growths [1–3].

In addition to clinical factors such as age, smoking history, gender, lesion location [4], and radiological nodule characteristics such as texture, size/growth rate [5, 6], and positron emission tomography standard uptake values [7], surface morphology and more generally shape have been identified as important independent predictors of malignant potential for solid lung nodules observed on CT [8–13]. Additionally, shape is useful for lung nodule detection and segmentation from blood vessels and chest wall [14–18], and has also been shown to affect nodule size calculation variability [19, 20]. Shape is also well known for helping to predict the behavior of breast lesions seen on mammography, ultrasound, and MRI [21–23].

There exist countless metrics of shape, the most popular falling into the categories of statistical moments, Fourier descriptors, or curvature measures [24]. These approaches, though broad, are united by their reliance on image geometric properties directly. Hence, they typically require significant intuition about which features will permit appropriate segmentation into the desired classes.

In the present work, I propose a novel approach to automated shape prediction, which I call Normal Mode Analysis Shape Detection (NMA-SD). The method connects an important tool from physical chemistry and structural biology (normal mode analysis) to image analysis by making the analogy of voxels to atoms in a molecule. Because they are not real atoms, and the nodule is not a real molecule, we denote the

J. N. Stember (✉)
Department of Radiology, Columbia University Medical Center,
180 Fort Washington Avenue, 3rd Floor Harkness Pavillion,
Room 313, New York, NY 10032, USA
e-mail: joestember@gmail.com

voxels as pseudoatoms and the nodules/shapes as pseudomolecules. We then apply normal mode analysis (NMA) to the pseudomolecules to yield the major motions that the nodule would undergo were it a molecule. By analyzing these motions, we can work backward to predict the nodule's shape, since this shape determines the aforementioned motions.

Such an approach has the advantage of enabling users to visualize the object's internal movements and develop an intuition for which motions are important, and which geometric features give rise to them. In this manner, it can help to identify appropriate classification features to distinguish among classes of interest.

As a proof-of-principle, I apply NMA to simulated nodules represented by pseudomolecules. The three shape types studied represent the broadest classifications of lung nodule morphology: smooth—which subdivides into circular and elliptical, and irregular, characterized chiefly by the presence of surface lobulations and spiculations [8].

Methods

We begin by generating $N_{\text{draw}}=10$ training and testing nodules for each of shape types circular, elliptical, and irregular. Matlab's *imfreehand* function enables tracing out approximate shapes from exact circles/ellipses. These shapes, within a class, are qualitatively similar to one another but not equivalent, due to the expected variations of drawing said shapes (Fig. 1). In the case of irregular morphology, we have generated nodules with different numbers and sizes of lobulations/spiculations. All shapes are then rotated by random angles. The aforementioned process yields a set of 30 training and 30 testing nodules.

Binary matrices represent the nodules, and we may assign their nonzero interior and surface pixels (*row*, *column*) positions to a corresponding set of (*x*, *y*) coordinates. We situate nodule image matrices in 3-dimensional space by assigning *z* values identically to zero.

At this point we recast the notion of pixels/voxels in a binary nodule image matrix into that of “atoms” in a “molecule”, or as we will denote them pseudoatoms in a pseudomolecule. We seek to envision the pseudomolecule's natural internal motions, consisting largely of the vibrations of pseudobonds that connect adjacent pseudoatoms. In order to do so, we employ NMA, a fundamental approach for predicting molecular motion.

The normal workflow of NMA is using a molecular structure to predict important motions. In our case, the nodule is not expected to actually move in any significant way. Hence, the movements predicted are not themselves biologically significant. However, they permit us to identify important aspects of

the underlying shape, since it is this shape that gave rise to the predicted motions.

Briefly, given a pseudomolecule containing N pseudoatoms, there are $3N$ so-called degrees of freedom, representing the fact that each pseudoatom can move in three directions. Denoting the set of these degrees of freedom by $\{r_i\}_{i=1}^{3N}$, it can be shown that for small displacements of the r_i 's from their equilibrium (lowest energy) positions $\{r_i^0\}_{i=1}^{3N}$, the expected energy function is quadratic in form. When this is true, it can be shown that the pseudoatoms move according to the equation

$$r_i(t) \propto r_i^0 + \sum_{k=1}^{3N} a_{ik} \cos(\omega_k t), \quad (1)$$

where ω_k is the angular frequency of normal mode k , and a_{ik} is degree of freedom i 's contribution to the motion of normal mode k [25]. We may calculate $a_k = \{a_{ik}\}_{i=1}^{3N}$ as the k^{th} eigenvector of the Hessian matrix, which has elements $H_{ij} \equiv \left(\frac{\partial^2 V}{\partial r_i \partial r_j} \right)_0$, where $V = V(\{r_i\}_{i=1}^{3N})$ is the pseudomolecule's potential energy and the zero subscript denotes definition about the equilibrium pseudomolecular configuration.

Following the Tirion model [26], for small displacements, we may estimate the terms of the potential energy V_{ij} by pairwise Hookean potentials between connected pseudoatoms:

$$V_{ij} \propto \left(\|r_{ij}\| - \|r_{ij}^0\| \right)^2, \quad (2)$$

where $\|r_{ij}\| = \|r_i - r_j\|$ is the distance between connected pseudoatom pairs i and j , and $\|r_{ij}^0\|$ is the equilibrium value of this distance. The full potential energy in the Elastic Network Model Normal Mode Analysis formulation (ENM-NMA) is then given by

$$V = \sum_{i < j} V_{ij} D_{ij}, \quad (3)$$

where $D_{ij} = \Phi(D - \|r_{ij}^0\|)$, Φ being the Heaviside step function and D a distance cutoff below which each pair of pseudoatoms is connected by a pseudobond. D is chosen to reflect the level of pseudomolecular detail that should translate into the motion. It is typically chosen so as to connect adjacent pseudoatoms while leaving out long-range pseudobonds that might artificially constrain the motion [26, 27]. I have chosen $D=6$ to suit the system studied here.

I have implemented the above approach via the program Modehunter (<http://modehunter.biomachina.org>) [28]. Doing so yields $3N$ normal modes of motion. However, three of these are the trivial center-of-mass translations, while three others are trivial center-of-mass rotations, leaving $3N-6$ non-trivial modes that represent internal vibrational movement. These modes appear as output in order of decreasing normal mode frequency, so that the 7th (first nontrivial) mode

Fig. 1 Representative nodules, training set. **a–c** are from the circular class, **d–f** are from the elliptical class, and **g–i** are from the irregular class

corresponds to the dominant, highest amplitude/lowest-frequency internal motion. Each subsequent mode represents increasingly high frequency, lower amplitude, finer movements that factor (to a decreasing degree) into the overall predicted molecular motion.

Before proceeding with NMA, we confer upon nodule pseudomolecules nonzero thickness by stacking together nodule image copies above and below at $z=+1, -1$, respectively, thereby yielding three-dimensional nodule pseudomolecules with widths of three pseudoatoms. Doing so provides a degree of steric hindrance to motions along the z -axis (and thus prevents unbounded motions along z) while still reflecting the overall shape of the two-dimensional nodular slices.

The intuitive idea behind our approach is that different structures should lead to different motions because of their varying shapes/pseudoatom connectivities. One would expect smaller differences in structure (such as the lobulations/spiculations in irregular nodules) to manifest at the higher-frequency modes. At these modes, the movements of smaller structural components factor more than the global dominant motions of the lower frequency modes.

The circular and elliptical nodules represent two ends of a spectrum. Due to their geometry, the circular nodules vibrate predominantly along the z -direction due to the extensive steric hindrance that motion along the xy plane would entail (Fig. 2). The long, slender elliptical nodules have less steric impairment along xy due to their rod-like shape and are thus able to bend about the z -axis in the xy plane and exhibit more complex twisting motions with large xy components at many of

the higher modes (Fig. 3). This points to a simple way to distinguish between the two shapes, namely the degree to which they sample the xy plane.

The irregular shapes are a kind of intermediate, with a compact, circular core, yet featuring lobulations/spiculations that protrude outward. The latter peninsulas can be thought of as small ellipses that can swing in the xy plane, thereby imparting an intermediate predilection for motion along said plane (Fig. 4). As suggested above, this peninsular motion manifests at the intermediate/higher frequencies of modes 11–25. We therefore expect the average, per-pseudoatom projection along xy for these intermediate normal modes to follow the trend: elliptical > irregular > circular.

In order to examine the aforementioned xy plane sampling more formally, we compute the projection of pseudoatom i along the xy plane via

$$\text{proj}_i = \left\| \mathbf{v}_i - (\mathbf{v}_i \cdot \hat{\mathbf{k}}) \hat{\mathbf{k}} \right\|, \quad (4)$$

where $\mathbf{v}_i$ is the normal mode vector for pseudoatom i and $\hat{\mathbf{k}}$ is the unit normal vector to the xy plane, i.e., in the z -direction [29]. Then to obtain the summed projection along xy scaled by pseudomolecular size, we add up the projections for every pseudoatom in the pseudomolecule and divide by the number of pseudoatoms:

$$\text{proj} = \frac{1}{N} \sum_{i=1}^N \text{proj}_i. \quad (5)$$

Fig. 2 Circular nodule pseudomolecular normal mode 12, testing set. Note that the motion is predominantly along the *z*-axis, with little projection onto the *xy* plane

Finally, we compute an average projection $\langle \text{proj} \rangle$ over the moderately high-frequency modes 11–25. Doing so for every pseudomolecule, we observe the expected trend mentioned above (Fig. 5).

Results

I have constructed a classifier to predict shape type based on the binary nodule image. Using a k-nearest neighbor classifier with neighborhood size of $N_{\text{draw}}=10$, the model correctly

predicts the shape of 29 testing set nodules, for an accuracy of roughly 96.7 %.

Conclusions

The present work shows that it is possible to classify shapes based on the predicted motion of the corresponding pseudomolecules. It should be noted that lung nodules are not expected to actually move significantly, so that their pseudomolecular motion does not directly predict anything biologically or medically. As shown here, however,

Fig. 3 Elliptical nodule pseudomolecular normal mode 12, testing set. Note the large swinging movements that contribute to a significant overall projection onto *xy*

Fig. 4 Irregular nodule pseudomolecular normal mode 11, training set. Note the swinging movements about two of the nodule's protrusions that contribute to the shape's intermediate xy projection

movements permit inference of nodule shape, an important predictor of malignant potential.

The present work examines three important and general shapes that are of particular relevance in the study of solid lung nodules: circular, elliptical, and irregular. However, the principles illustrated should apply to any of a range of two- and three-dimensional shapes.

In the present work, I have applied NMA-SD to two-dimensional nodule slices, which is particularly useful because radiologists typically scroll through two-dimensional

axial planes for chest CT. Most of the quantitative tools that radiologists use to inform their interpretation of such images (for instance measuring Hounsfield Units) are applied to the two-dimensional image slices. Though I have applied NMA-SD to two-dimensional nodule slices, these slices are of nonzero thickness, so that they exist and move in three-dimensional space, as they need to in order for NMA to work properly. Thus, the method is directly applicable to three-dimensional image representations.

Fig. 5 Average (normal modes 11–25) per-pseudoatom projections along xy, testing set. These clearly follow the expected trend of elliptical (*filled dots*) > irregular (*crosses*) > circular (*filled squares*)

In the work presented here, a very simple measure of motion successfully partitions the three shape types into classes. However, one can imagine using many other metrics, such as the distribution of motion vectors and measures of particular types of motion (e.g., bending or twisting of particular image components) that are characteristic to a given shape or surface feature. Multiple measures used concurrently can further distinguish shape classes.

It is well known from structural biology that the very lowest-frequency nontrivial modes (typically around modes 7–10) contribute the most to the native motion by reflecting global geometry and connectivity. Higher-frequency modes reflect smaller differences in structure than their low-frequency counterparts. Analyzing the intermediate and higher frequencies is therefore useful for exploring more modest disparities in shape, which can be important radiologically.

Thus, normal modes examined should be high enough to study the particular differences in shape that are of interest. On the other hand, the modes should not be too high, in which case they would reflect size features smaller than those being studied. The window of appropriate modes needs to be adjusted depending on the system one is studying.

NMA-SD is invariant to rotation and translation, given that only the internal pseudomolecular geometry and pseudobonding factor into predicted motions. It is also scale invariant, because the pseudobond cutoff distance can be adjusted to mirror the overall scale.

An important limitation of the method is its time of computation. This will in general depend on a host of factors including the particular implementation of NMA that one is using. For the system studied here, NMA wall-clock time ranges from around 25–65 s for a nodule. For the simple classification task of the present work, many shape descriptors would perform the task in a much shorter amount of time.

A significant advantage of NMA-SD is the ability to visualize normal modes and thereby gain insight into meaningful motions, the latter reflecting key structural themes. Such insights can guide the selection of appropriate features that will allow effective partitioning into classes of interest. Widely available programs such as Modehunter with VMD, CHARMM, and AMBER [28, 30–32] greatly enhance visualization via normal mode animations. These animations can allow users to visualize pseudomolecular movements, which provide an intuitive route to identifying key structural elements that distinguish different shapes. The insights thus obtained can also help to uncover appropriate features in faster shape classifiers.

Future work will involve computing normal modes and finding appropriate features of the resulting motion for other systems of interest. One potentially useful tool may be the Bend-Twist-Stretch method, which is able to animate any given connected graph [28].

References

1. El-Baz A, Nitzken M, Elnakib A, et al: 3D shape analysis for early diagnosis of malignant lung nodules. Medical image computing and computer-assisted intervention: MICCAI International Conference on Medical Image Computing and Computer-Assisted Intervention. 14(Pt 3):175–182, 2011
2. Webb WRHC: Thoracic Imaging: Pulmonary and Cardiovascular Radiology. 2nd edition. Wolters Kluwer, 2011
3. Jacobson FL, Judy PF, Feldman U, Seltzer SE: Perceived features reported as nodules: interpretation of spiral chest CT scans. Acad Radiol 7(2):77–82, 2000
4. McWilliams A, Tammemagi MC, Mayo JR, et al: Probability of cancer in pulmonary nodules detected on first screening CT. N Engl J Med 369(10):910–919, 2013
5. Ko JP, Rusinek H, Jacobs EL, et al: Small pulmonary nodules: volume measurement at chest CT—phantom study. Radiology 228(3):864–870, 2003
6. Ko JP, Berman EJ, Kaur M, et al: Pulmonary Nodules: growth rate assessment in patients by using serial CT and three-dimensional volumetry. Radiology 262(2):662–671, 2012
7. Nie Y, Li Q, Li F, Pu Y, Appelbaum D, Doi K: Integrating PET and CT information to improve diagnostic accuracy for lung nodules: A semiautomatic computer-aided method. J Nucl Med Off Publ Soc Nucl Med 47(7):1075–1080, 2006
8. Iwano S, Nakamura T, Kamioka Y, Ishigaki T: Computer-aided diagnosis: a shape classification of pulmonary nodules imaged by high-resolution CT. Comput Med Imaging Graph Off J Comput Med Imaging Soc 29(7):565–570, 2005
9. McNitt-Gray MF, Wyckoff N, Sayre JW, Goldin JG, Aberle DR: The effects of co-occurrence matrix based texture parameters on the classification of solitary pulmonary nodules imaged on computed tomography. Comput Med Imaging Graph Off J Comput Med Imaging Soc 23(6):339–348, 1999
10. Stember JN, Ko JP, Naidich DP, Kaur M, Rusinek H: The self-overlap method for assessment of lung nodule morphology in chest CT. J Digit Imaging 26(2):239–247, 2013
11. Shah SK, McNitt-Gray MF, Rogers SR, et al: Computer aided characterization of the solitary pulmonary nodule using volumetric and contrast enhancement features. Acad Radiol 12(10):1310–1319, 2005
12. Ruppert AM, Lerolle U, Carette MF, et al: Coexisting pulmonary nodules in operable lung cancer: prevalence and probability of malignancy. Lung Cancer 74(2):233–238, 2011
13. Li F, Sone S, Abe H, Macmahon H, Doi K: Malignant versus benign nodules at CT screening for lung cancer: comparison of thin-section CT findings. Radiology 233(3):793–798, 2004
14. Choi WJ, Choi TS: Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor. Comput Methods Prog Biomed 113(1):37–54, 2014
15. Magalhaes Barros Netto S, Correa Silva A, Acatauassu Nunes R, Gattass M: Automatic segmentation of lung nodules with growing neural gas and support vector machine. Comput Biol Med 42(11):1110–1121, 2012
16. Diciotti S, Lombardo S, Falchini M, Picozzi G, Mascalchi M: Automated segmentation refinement of small lung nodules in CT scans by local shape analysis. IEEE Trans BioMed Eng 58(12):3418–3428, 2011
17. Murphy K, van Ginneken B, Schilham AM, de Hoop BJ, Gietema HA, Prokop M: A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification. Med Image Anal 13(5):757–770, 2009
18. Ye X, Lin X, Dehmshki J, Slabaugh G, Beddoe G: Shape-based computer-aided detection of lung nodules in thoracic CT images. IEEE Trans BioMed Eng 56(7):1810–1820, 2009

19. Petrou M, Quint LE, Nan B, Baker LH: Pulmonary nodule volumetric measurement variability as a function of CT slice thickness and nodule morphology. *AJR Am J Roentgenol* 188(2):306–312, 2007
20. Gietema HA, Wang Y, Xu D, et al: Pulmonary nodules detected at lung cancer screening: interobserver variability of semiautomated volume measurements. *Radiology* 241(1):251–257, 2006
21. Huang YL, Chen DR, Jiang YR, Kuo SJ, Wu HK, Moon WK: Computer-aided diagnosis using morphological features for classifying breast lesions on ultrasound. *Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol* 32(4):565–572, 2008
22. Moon WK, Lo CM, Chang JM, Huang CS, Chen JH, Chang RF: Quantitative ultrasound analysis for classification of BI-RADS category 3 breast masses. *J Digit Imaging* 26(6):1091–1098, 2013
23. Nie K, Chen JH, Yu HJ, Chu Y, Nalcioğlu O, Su MY: Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI. *Acad Radiol* 15(12):1513–1525, 2008
24. Costa LF, Cesar RM: Shape analysis and classification : theory and practice. CRC Press, Boca Raton, 2001
25. Cui Q, Bahar I: Normal mode analysis : theory and applications to biological and chemical systems. Chapman & Hall/CRC, Boca Raton, 2006
26. Tirion MM: Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys Rev Lett* 77(9):1905–1908, 1996
27. Chacon P, Tama F, Wriggers W: Mega-Dalton biomolecular motion captured from electron microscopy reconstructions. *J Mol Biol* 326(2):485–492, 2003
28. Stember JN, Wriggers W: Bend-twist-stretch model for coarse elastic network simulation of biomolecular motion. *J Chem Phys* 131(7):074112, 2009
29. Schneider PJ, Eberly DH: Geometric tools for computer graphics. Morgan Kaufmann Publishers, Amsterdam, 2003
30. Humphrey W, Dalke A, Schulten K: VMD: visual molecular dynamics. *J Mol Graph* 14(1):33–38, 1996. 27–8
31. Brooks BR, Brooks 3rd, CL, Mackerell Jr, AD, et al: CHARMM: the biomolecular simulation program. *J Comput Chem* 30(10):1545–1614, 2009
32. Case DA, Cheatham 3rd, TE, Darden T, et al: The Amber biomolecular simulation programs. *J Comput Chem* 26(16):1668–1688, 2005