Published in final edited form as:

Neuroimage. 2008 February 15; 39(4): 1721–1730. doi:10.1016/j.neuroimage.2007.10.043.

Measuring Structural Complexity in Brain Images

Karl Young and Norbert Schuff

Department of Radiology, University of California San Francisco and Center For Imaging of Neurodegenerative Diseases, Department of Veterans Affairs Medical Center, San Francisco, CA 94121

Abstract

An information theory based formalism for medical image analysis proposed in [Young, K., Chen, Y., Kornak, J., Matson G. B., Schuff, N., 2005. Summarizing Complexity in High Dimensions, *Phys. Rev. Lett.* 94 098701-1] is described and used to estimate image complexity measures as a means of generating interpretable summary information. An analysis of anatomical brain MRI data exhibiting cortical thinning, currently considered to be a sensitive early biomarker for neurodegenerative diseases, is used to illustrate the method. Though requiring no previous assumptions about the detailed shape of the cortex or other brain structures, the method performed comparably (sensitivity = 0.91) to direct cortical thickness estimation techniques (sensitivity = 0.93) at separating populations in a data set designed specifically to test the cortical thickness estimation algorithms. The results illustrate that the complexity estimation method, though general, is capable of providing interpretable diagnostic information.

INTRODUCTION

Modern neuroimaging increasingly requires sophisticated tools for extracting interpretable summary information from the growing volumes of data, such as magnetic resonance imaging (MRI) with various contrasts [1,2,3]. Although supervised learning and clustering methods [4] have been used to reduce the dimensionality of the data, the results often lack a framework for interpretation and are hard to generalize [5]. Another critical issue in image analysis is that the spatial characteristics can vary from point to point, that is they are non-stationary. Although popular analysis techniques that assume stationarity, such as random field theory methods [6], have effective ways of addressing non-stationarity, they risk underutilizing some of the information contained in spatial patterns. Methods that do not rely on stationarity, such as texture analysis [7,8] have proved fruitful in particular cases of image classification. But they typically do not generalize well, for example to different modalities or higher feature space dimensions. Furthermore, texture analysis often fails to provide insight into the underlying structural changes [8]. There is therefore need for a method that maximally extracts information from the available data, and does so in a way that is interpretable for purposes such as diagnostic classification. The main objective of this paper is to show that information theory based complexity estimation measures can be utilized for deriving interpretable summary information from structural MRI data of the brain.

Corresponding Author: Karl Young, Center for Imaging of Neurodegenerative Diseases, VA Medical Center, 4150 Clement Street, San Francisco, CA 94121, Phone: (415) 221-4810 ext. 3114, FAX (415) 668-2864, karl.young@ucsf.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

An important aspect of the information theoretic complexity estimation methods presented in this paper is that they provide summary information in the form of quantitative measures that can be generalized to multivariate studies. The fundamental hypothesis is that the complexity of spatial patterns in neuroimages (such as the convoluted spatial distribution of human cortex evident in MRI) can effectively be captured by complexity measures that correspond well with visual impressions of image complexity but generalize easily to use with any number of combined imaging modalities.

A basic outline of the procedure described in this paper for estimating complexity measures from single or multiple images is as follows. If there are multiple images obtained for each subject the images are first coregistered and smoothed or interpolated so that all transformed images are at the same resolution. Segmentation is then performed in the resulting feature space; if there are N images the feature space is defined as the N -dimensional space of values from all voxels in the images without regard to spatial position. For example for a single image the feature space can be represented as an intensity histogram. Next a pair of spatial templates is chosen. For example a typical spatial template is a box of a given dimension in units of voxels. A simple case of a spatial template pattern, used in this paper, is a linear set of voxels of a given length. This template structure will be referred to as a linear template of length L . Two templates are chosen so as to assess the conditional probabilities of image patterns over one template given the pattern observed in the other. These probabilities are estimated by parsing, or moving the templates over the entire image or region of the image being studied and accumulating a joint histogram of the patterns observed over the templates. While the general formalism and resulting algorithms place no restriction on the form of the templates or their spatial relationship, results generated using simple, contiguous linear template pairs are easy to interpret and were used in this paper. Complexity measures, defined below, are estimated from the joint histogram. These complexity measures then serve as summary values for the image or image region under consideration.

In this paper complexity estimates are used for studying population differences in an experimental set of 3D structural brain MRI data for which cortical thinning was artificially applied [9] to simulate effects that are commonly seen in Alzheimer's disease and other neurodegenerative diseases. In this case the original image space is the univariate intensity in the T1 image and the segmented image is that obtained by applying a commonly used tissue segmentation algorithm. This example was chosen to demonstrate in particular that: 1) use of complexity measure estimation is directly comparable with state-of-the-art image analysis methods for detection of neurodegenerative diseases, such as Alzheimer's disease and frontotemporal dementia [10,11,12,13,14,15,16] and 2) use of complexity measure estimation yields interpretable results of potential clinical relevance, despite the generality of the algorithms.

METHODS

Theory

The complexity estimation methods used in this study were introduced in [17] and subsequent papers [18,19,20] for time series analysis. The fundamental question addressed in [17], was how much model complexity is required to optimally predict future values of a time series from past values. In addition a framework was provided for building the optimal model in the sense of being the minimally complex model required for reliable predictions. Heuristic arguments and examples provided in [21] showed that only slight modifications were required to generalize the formalism for analysis of spatial data and in particular medical image data.

Critical to the definition of complexity is the notion of "state" which provides the ability to predict observed values in a time series or image. Simply put, the complexity of the set of states

required to describe a particular time series or image, is an indication of the complexity of the system that generated the time series or image. As described in [18] to accurately and rigorously characterize the underlying complexity of a system, the set of states must in fact constitute a minimal set of optimally predictive states. How those criteria are defined and satisfied by the construction described in this paper is described in [18].

The simplest notion of complexity that arises from the above considerations is a count of the number of states required for making optimal predictions. Since enumerated states can occur with varying frequencies during the course of observations, introducing the notion of state probability is natural. Utilizing Shannon's original criteria for information [22], the simplest

definition of an additive quantity associated with the set of states is $-\sum_{i=1}^S p_i \ln p_i$, where p_i denotes the estimated probability of the i^{th} state and the sum is over the S states. Complexity can then be described as an extensive quantity (i.e. a quantity that scales with size) associated with the system it describes. For equally probable states this definition yields $\ln S$ as a measure of complexity.

This notion of complexity, based on considerations of optimal prediction, is very different from the traditional notion of Kolmogorov complexity [23], which quantifies series of random values as the most complex, based on considerations of incompressibility of a sequence. Here sequence is interpreted as a "program" in the context of computation, and data in the context of data analysis. We will show that both notions of complexity provide important and complimentary measures for characterizing structure in images. In the following, the optimal prediction based definition of complexity is the statistical complexity (SC) and the incompressibility based definition of complexity is entropy (H), since the Kolmogorov complexity corresponds, in the case of data analysis, to what physicists typically refer to as entropy [23].

A brief description of the definitions used here to make complexity estimates from univariate or multivariate images follows.

Definitions

The components required for defining SC and H are an index set I representing a set of spatial or space/time coordinates and a feature space F defined over I , representing some set of variables defined at each index. The feature space used in this paper consists of the values at each voxel of a set of coregistered images in a standardized space [9], in particular the 1 dimensional space of intensity values from T1 images.

Definition of the complexity measures requires the conditional probability

$$P(Z=M(F^{T^1})|Z'=M(F^{T^2}))=\frac{P(Z=M(F^{T^1}), Z'=M(F^{T^2}))}{P(Z'=M(F^{T^2}))} \quad (1)$$

defined in terms of the joint empirical distribution over observed patterns

$$P(Z=M(F^{T^1}), Z'=M(F^{T^2}))=\frac{1}{N} \sum_j \sum_i \delta(Z=M(F^{T^1}_j), Z'=M(F^{T^2}_i)) \quad (2)$$

where the template pair, T_j^1 and T_i^2 , are ordered sets of indices forming distinct but possibly overlapping regions of voxels, in the underlying index set I , at positions j and i respectively. N is the total number of template pairs observed over the whole image or region. F^{T^k} is a product feature space, which is the product space of values in the image(s) at voxels in the template T^k . The mapping $M: F^{T^k} \rightarrow \mathbb{Z}$, maps the set of all patterns observed in the image or region over the product feature space F^{T^k} , to the integers. That is M indexes the observed patterns, assuming that the set of patterns form a discrete set. The sums are taken over all particular instances, T_i^k , that is instances of template k “located” at index i . The functional $\delta(\cdot)$ is the indicator “function”, yielding 1 if a particular pair of patterns is observed, 0 otherwise. Thus $P(Z = M(F^{T^1}) \mid Z' = M(F^{T^2}))$ effectively constitutes a 2D histogram of counts of feature space patterns observed when parsing over the index space with the templates T_j^1 and T_i^2 . Further restrictions can be imposed such as requiring that T^1 and T^2 be non-overlapping and/or contiguous. In the analysis of cortical thinning presented below T^1 and T^2 are linear sets of contiguous voxels of a given length L with the two templates required to be contiguous. Hence the analysis effectively studies local correlation structure at scale L . In the following we will assume the mapping M exists and has been made and hence simplify the notation using the conversion $Z = M(F^{T^k}) \rightarrow Z$. The marginal distribution is then defined as:

$$P(Z') = \frac{1}{N} \sum_Z \delta(Z, Z') \quad (3)$$

To estimate SC the set of optimally predictive states must be determined from the images as described in [17,19,21]. For the purposes of this paper there is no need to provide an explicit representation of the states described in [17,19].

Determination of states—Initially the image or multiple images are segmented using some form of clustering. For this paper this step is simply tissue segmentation of the univariate T1 image. Better segmentation algorithms provide some benefit in obtaining better complexity estimates but complexity estimation as a whole, given the global nature of the analysis, can be viewed as less sensitive to the accuracy of segmentation algorithms than techniques that rely on accurate “local” information such as cortical thickness estimation.

The next step in the algorithm is to choose templates and parse the image using those templates. That is the templates are moved over the image and the image values over the templates are recorded as counts in a joint histogram. The number of joint histogram bins $s^{v_1} \times s^{v_2}$ is determined by the number of segmentation values, s , in the image and the number of voxels, v_1 and v_2 in the templates. s^{v_1} is the number of possible patterns that can be observed over template 1 and s^{v_2} is the number of possible patterns that can be observed over template 2. For this paper the two template structures are identical, simple linear sequences of voxels, and s^{v_1} and s^{v_2} represent the same number.

From the joint histogram the conditional histograms defined in equation (1) are obtained. The conditional histograms are then grouped, based on a measure of similarity for probability distributions; there are a number of choices for similarity measure between distributions. In this paper a hard clustering algorithm, PAM [27], is used with the Euclidean distance between bin counts as the similarity measure. Note that the grouping of conditional histograms into states is the step that specifically characterizes structure in the image and distinguishes the complexity based method from simply estimating entropies from the joint histogram as a co-occurrence matrix [28]. This step is also what distinguishes SC from the joint entropy though in the special case where there is a one to one correspondence between joint histograms and

states, SC reduces to the joint entropy. This can occur when there are a small number of joint histograms that are too distinct to be clustered into states.

The ability to specify arbitrary template structures also distinguishes the complexity estimation methods from standard co-occurrence analysis. In addition, while the joint histogram, analogous to a co-occurrence matrix, is a convenient representation of the results of the image parsing step, the underlying construction is based on a rigorously defined and general set of graph theoretic methods described in [29].

Statistical Complexity (SC)—Given the set of states s , accumulated as just described, SC can then be defined as:

$$SC = - \sum_s \left(\sum_{Z, Z' \in s} P(Z|Z') \right) \log_2 \left(\sum_{Z, Z' \in s} P(Z|Z') \right) \quad (4)$$

This quantifies the information contained in the distribution of observed patterns, conditioned and summed over all states, s . Note that as mentioned above, via conditioning over the states s , SC is distinct from the conditional entropy of the histogram and measures something quite different. For example when the set of N conditional histograms making up the joint histogram are identical the joint entropy would equal N times the entropy of one of the conditional histograms. But SC would equal zero as there is only a single state.

Though not explicitly represented in equation (4) SC is dependent on the particular templates T^1 and T^2 over which the patterns are observed and implicitly on the scale of the template patterns. However, as will be demonstrated below, as long as the choice of template patterns is consistently applied, the results can be used to detect systematic differences between images.

Entropy (H)—The standard measure of the number and distribution of observed patterns produced by a system is the entropy

$$H = - \sum_{Z', Z} P(Z, Z') \log_2 P(Z, Z') \quad (5)$$

where the joint probabilities were defined in equation (2). As discussed above the entropy corresponds to the traditional definition of Kolmogorov complexity and a complimentary definition of complexity to SC . The entropy in the current case is over the joint distribution defined in equation (2), accumulated over the underlying template pairs and summed over the indices of the observed patterns.

Excess Entropy (EE)—Convergence rate estimates of H have proved useful in time series analysis [18]. A problem arises in trying to simply generalize convergence rate estimates to image analysis due to the fact that stationarity of spatial characteristics cannot generally be assumed, as mentioned in the introduction. Another problem is that there is no clear order of directions in which to take appropriate limits. In particular asymptotic values typically depend on that choice of order [24,25]. In that situation, the variance of H across a range of scales may cause complications for characterizing entropy convergence by EE . However this variance, if systematic, may indicate interesting scaling properties in the images as illustrated by results of measuring the fractal dimension of various brain structures [30]. EE was defined and initially discussed in [31] in the case of index sets with dimension higher than 1. We first define the

metric entropy as the limit of entropy per unit volume as the volume is taken to infinity, initially without specifying explicitly how the volume is to be taken to infinity:

$$h_{\mu} = \lim_{V \rightarrow \infty} \frac{H}{V} \quad (6)$$

EE is then defined as:

$$EE = \sum_{V \sim 0}^{\infty} \left(\frac{H}{V} - h_{\mu} \right) \quad (7)$$

where V is a measure of the volume of the templates in index space. Despite the subtleties involved with estimating EE in dimensions higher than 1, any particular choice, consistently applied, results in a useful method for comparing structure in images. In particular note that if the entropy per unit volume converges quickly to h_{μ} as a function of scale then EE is small indicating a lack of large scale structures in the image.

EE provides an estimate of the scaling properties in image data that compliments information obtained using SC and H . Using cortical thinning as an example, we demonstrate below that EE combined with estimates of H and SC at particular scales does provide a comprehensive structural description of the images. Specifically H provides a measure of the number and distribution of structures observed at a given template scale, SC provides a measure of the complexity of the spatial correlation of those structures at a given template scale, and EE provides a measure of the variation in the number of observed structures as a function of template scale. In the case of specific focal cortical thinning examined below, these quantities exhibit characteristic behavior that will be discussed.

Implementation

The methods proposed in this study and described above are similar to those described in [17,21] and are represented schematically in a simple application to human brain structural MR data in figure (1). For the particular case of MRI, the steps for obtaining sets of states and calculating SC , H , and EE are as follows:

1) Feature space reduction—For the analysis in this paper, feature space reduction, had already been performed given that the complexity estimation steps were applied to the segmented T1 image data described in [9]. That is, the feature space consists of a single value, the intensity of a T1-weighted MR brain image, and the feature space reduction corresponds to standard tissue segmentation of the brain image into 3 classes, gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF).

2) Parsing—The index space is parsed according to the specified templates. The observed patterns over all instances of the templates in the index space are then catalogued via the joint histogram. In this study parsing is done for the 3 dimensional index space in the cardinal directions (x, y, z), and templates consist of linear sequences of voxels in those directions.

To obtain the list of distinct states, as describe above, a robust hard clustering algorithm (PAM [27] in the current implementation) is used to group the conditional distributions. There are various similarity measures that could be used to group the conditional histograms into states but use of PAM for clustering based on Euclidean distance between histogram bin counts proved to be relatively robust in being transparent to numerical problems resulting from histogram bins with zero counts.

3) Complexity estimation—The total probabilities of the states, that is the normalized sums of the histogram counts in the conditional histograms in that cluster, are obtained and used to estimate SC using equation (4). H is estimated over all histogram bins treated independently using equation (5).

The software implementation of the above methods is an open source package written in Python [32] and SciPy [33], and uses the Rpy [34] package to provide access to the statistical and graphical capabilities of the R statistical language [35] and supplemental libraries. The cluster [27] and e1071 [36] R packages were used for clustering and the AnalyzeFMRI [37] package for MR image processing. Image analysis was performed using this package on a 46 processor Beowulf cluster using the PyPAR [38] Python wrapper for the message passing interface MPI.

Simulations

Automated Detection and Classification of Changes in Cortical Thickness—To test the procedures on a well defined, hypothesized condition of brain abnormality (cortical thinning) on MRI, the analysis was performed using the ICBM data set described in [39], which consists of experimental T1-weighted brain MRI data from 50 healthy subjects, segmented into GM, WM and CSF. 25 of the images had the cortical surface in the right superior temporal gyrus (RSTG) artificially thinned by an average of 0.5 mm which was less than the actual 1 mm resolution, of the segmented images. The data set was originally generated to study the efficacy of various cortical thickness measurements described in [9]. Use of this data set for the purposes of the current study was important in demonstrating the robustness of the estimated complexity measures to normal population variations in brain structure. The first step in the analysis was histogram generation for all length L sequences of symbols representing segmented values. The feature space reduction had already been performed in this case by the segmentation of the initial T1 weighted images into a set of categorical variables (GM, WM, and CSF), with one at each point over the index space. Since the image resolution was 1 mm and average thinning was 0.5 mm, the minimal sequence length of 2 mm turned out to be optimal for histogram generation in terms of capturing the effects of cortical thinning and other effects on this scale. This was determined a posteriori by automatically accumulating histograms over scales of 2 mm to 12 mm and not by any a priori assumptions of structure at that scale. The methods are most sensitive at whatever scales important structure occurs and this is determined automatically. Calculation of EE requires an extrapolation to length 0 from longer lengths, as described in [31]. Histograms for linear templates of length up to 12 mm ($L = 12$) were used to estimate EE as H/L changed by less the 5 % of its initial value between $L = 11$ and $L = 12$. The estimation of H and SC for histograms generated by parsing with the $L = 2$ (2 mm) templates were automatically found to be most sensitive and results for those values are reported.

Complexity estimates were obtained for successively refined regions of: whole brain (WB), temporal lobe (TL), superior temporal gyrus (STG), and right and left superior temporal gyrus (RSTG, LSTG). The regions were delineated using a 2 mm resolution labeled version of the MNI brain region atlas [40] superimposed on the images (1 mm resolution) which had been coregistered and warped to the same space as the atlas [9] so some statistical variation due to imprecise region delineation is expected. It should be noted that the reason for using atlas based region delineation was primarily to simplify the analysis for different brain regions and to efficiently repeat the analysis for arbitrary region as well as providing for direct comparison of the sensitivity of the complexity based analysis with that of cortical thickness estimation. Of course, to meaningfully compare regional estimates across an atlas, mapping is required and in that case utilization of region labels associated with a given atlas can serve to increase the efficiency of the method. But the complexity estimation method is perfectly general and does not itself rely on atlas mapping, or initial tissue segmentation. For example if a global

effect were suspected, complexity estimation could be performed directly on subjects with no initial tissue segmentation (clustering methods would be applied directly to the possibly multimodal image data) or atlas mapping.

In addition, to normalize for population variations in brain structure, H and SC for each region were divided by the equivalent 12 mm quantity from the whole brain analysis. As EE effectively uses values over all scales within a region it was divided by a measure of tissue asymmetry ($(\text{total GM} - \text{total WM})/(\text{total GM} + \text{total WM})$) in the region for each subject to normalize for what was assumed to be non disease specific anatomical structure. To assess the significance for separating the groups (thinned and not thinned), a Wilks lambda test for a MANOVA analysis on H , SC , and EE was used. To test use of the combined complexity measures in a predictive model, linear discriminant analysis (LDA) was used. Welch two sample t-tests on the resulting first linear discriminant were performed to assess the efficacy of the LDA classification. Using any subset of H , SC , and EE for the analysis yields significance levels (p-values) larger by at least two orders of magnitude. Why these numbers were included is discussed below. This provides some indication that the summary information provided by the complexity estimates is at least to some degree independent. The sensitivity and specificity of detecting cortical thinning were assessed using 10 times 10-fold cross validation for the LDA based classification results. These results are compared with the results generated by cortical thickness measurements and reported in [9].

RESULTS

As expected, the ability to separate the group with cortical thinning in the RSTG from the group with no thinning increased as the complexity estimation calculations were progressively constrained from WB to a region approximately encompassing the RSTG. Also as expected, an analysis only encompassing the LSTG, the only region containing no thinning at all, produced no separation between the groups. Figure (2) shows the values of H , SC , and EE in the RSTG for all 50 subjects, those for which the cortex was thinned represented by the open squares, those for which there was no thinning as closed circles. The figure shows a clear separation between the two groups. Furthermore, the scatter plots show that both SC and H have, on average, higher values when the cortex is thinned, consistent with the visual impression that regions of variable cortical thickness look more complex on MRI than those where cortical thickness remained constant.

Figure (3) shows the scatter plot of H , SC , and EE values in the LSTG, which was not thinned, for all 50 subjects, indicating no separation of the groups, as expected. The significance test results are summarized in Table I. Already at the level of the temporal lobe some ability to significantly separate the groups is apparent despite the fact that cortical thinning occurs only in the RSTG, a small fraction of the temporal lobe. The results become highly significant at the level of the superior temporal gyrus, which includes the RSTG, and finally for the RSTG itself. Even using maximally conservative Bonferroni multiple comparisons corrections for analysis of all possible brain regions labeled by the MNI atlas [40], on the order of 100 regions, the results for the RSTG are highly significant compared to an analysis limited to the LSTG. This was the reason for including the p-values to the level quoted. Also indicated in Table 1 is that LDA using the complexity estimates in the RSTG correctly identified cortical thinning with a sensitivity of 0.91 compared to the best sensitivity, 0.93, obtained from application of the suite of cortical thickness estimation algorithms tested in [9].

DISCUSSION

The sensitivity and specificity of the complexity estimation results indicate that it is capable of providing interpretable information of potential diagnostic value. While a number of

information theory based approaches have been applied in medical imaging, primarily for registration of images [41], the methods described in this paper represent a novel approach to summarizing structural information in images. Given the generality of the introduced measures SC , H , and EE , the method should be particularly useful for obtaining interpretable summary information from large, multivariate, medical imaging sets, for the purpose of diagnostic classification.

Recent brain MRI studies have suggested that thinning of the cortex is a hallmark of neurodegenerative diseases, such as Alzheimer's disease [10,11,12,13,14,15,16]. As such experimental MR images from "normal" brains from the ICBM data base with artificially induced cortical thinning provided an ideal test bed for demonstrating the utility of complexity estimation to detect MRI abnormalities of potential clinical relevance. Several methods have recently been reported for direct detection of cortical thinning [9,42,43]. These methods, though yielding promising results, depend on a number of parameters for spatial smoothing, mapping the cortical surface, and choosing a thickness metric [9], and thus are difficult to generalize to uses other than direct estimation of cortical thickness. In contrast, the complexity estimation method presented in this paper requires minimal intervention in terms of parameter choices. For example routine tissue segmentation methods were used and other than use of atlases for regional subject comparisons (used in any study of regional effects), no prior spatial structure was assumed.

The results using complexity estimation compared favorably with direct estimation of cortical thickness [9] though the data set was carefully constructed so that this was the only group effect present. In [9], 0.93 was the best sensitivity for group separation obtained by any method specifically designed for cortical thickness estimation, on the same data for which the complexity estimation method obtained a sensitivity of 0.91. In contrast SC , H , and EE should be more effective in quantifying combined variations in sulcal depth, cortical area, and cortical thickness in structural images. For example combined reductions in sulcal depth and cortical area with no change in cortical thickness would be expected to lead to image regions of greater homogeneity in terms of flatter areas of cortex. This would lead to an increase in H , and a decrease in SC and EE . The increase in H would result from an increased homogeneity in the joint histogram. The decrease in SC would result from a decrease in the number of predictive states required to describe the image. The reduction in EE would result from a faster convergence of H as a function of template scale. These changes would presumably be very difficult to detect with algorithms aimed at detecting only changes in cortical thickness.

Although the analysis illustrated that simpler methods, such as co-occurrence analysis, might have sufficed to separate the two groups in the simulated cortical thinning data, this was discovered in retrospect and as a result of the interpretation of the complexity estimation methods. There is no indication of how to generalize simpler methods, such as co-occurrence analysis, to the analysis of other disease specific effects. This is particularly true in the case of multimodal imaging data such as some combination of structural, spectroscopic, perfusion, diffusion MRI or MRI together with PET, SPECT. In contrast, the complexity estimation methods provide an explicit algorithm for generalization by utilizing SC , H , and EE for detecting changes in spatial distributions characterized by any set of coregistered image modalities, whether structural, functional or metabolic.

Application in the current study was restricted very specifically to structural images with simulated, focal cortical thinning to demonstrate the utility of the method in an easily interpretable context. As shown in the scatter plot in figure (2), both H and SC are generally larger for the subjects for which the cortex was thinned (open squares in the figure). Though this may seem counterintuitive it can be explained by noticing that the nearest neighbor tissue histograms ($L = 2$) generated from images in the superior temporal gyrus, are dominated by

two peaks in the conditional histogram. One peak represents contiguous gray matter voxels and the other, a smaller peak, represents contiguous white matter voxels. In the case of the simulated cortical thinning, implemented in [9], which selectively reduced gray matter, the asymmetry between gray and white matter peaks conditioned on white matter is smaller than the gray and white matter peaks conditioned on gray matter. Therefore in the regions of artificial cortical thinning the white matter peak grows and the gray matter peak shrinks making them closer in value, thus also making both the overall histogram and the conditional histograms more uniform. Hence, both the entropy of the overall histogram and the statistical complexity, increase.

As discussed above, when considering raw, non-segmented, image values a method of segmentation must be applied before the complexity estimates can be obtained. It is certainly the case that the choice of segmentation method, like choice of template structure, could affect the values of the estimated complexity measures obtained. But initial tests on simulated multimodal spectroscopic imaging data [21] using k-means, c-means fuzzy clustering, and a set of divisive and agglomerative clustering algorithms [27] for segmentation, indicates that the relative differences of the complexity estimates for a variety of image variations remain insensitive to which clustering method is used.

Some limitations of the current methods include the requirement of a choice of template structure, and clustering methods for generating the complexity estimates as described above. While the choice of simple linear template structure used for the analysis in this paper proved appropriate for sensitive diagnostic classification, it is not clear that this will always be the most appropriate choice. Exploration of various template structures in the context of multivariate image data is the subject of current work. The clustering method used for the analysis in this paper was chosen by default, i.e. it was the method used for tissue segmentation of the original data [9]. In multivariate applications of the methods, as discussed above and in more detail in [21] c-means fuzzy clustering is the method currently used and constitutes a reasonably robust choice. However limitations of c-means fuzzy clustering in medical imaging have been discussed and alternatives are currently being investigated [46]. More exploration of appropriate choices of clustering algorithm for multivariate studies is clearly required. In addition recent improvements in segmentation algorithms [47] and generation of high resolution, disease specific group atlases [48] should further improve the sensitivity of region specific versions of the complexity estimation analysis.

Acknowledgments

The authors thank George Bartzokis for an important suggestion that led to this study, John Kornak, Cosma Shalizi, and Brian Soher for helpful comments, Yue Chen for programming support, Alan Evans and Jason Lerch for generously providing their data, and Gerald Matson and Michael Weiner for generous support of this work primarily through NIH Grant No. 5RO1EB000766.

References

1. Johnson NA, Jahng GH, Weiner MW, Miller BL, Chui HC, Jagust WJ, Gorno-Tempini ML, Schuff N. Pattern of cerebral hypoperfusion in Alzheimer disease and mild cognitive impairment measured with arterial spin-labeling MR imaging: initial experience. *Radiology* 2005;234(3):851–859. [PubMed: 15734937]
2. Salat DH, Tuch DS, Hevelone ND, Fischl B, Corkin S, Rosas HD, Dale AM. Age-related changes in prefrontal white matter measured by diffusion tensor imaging. *Ann N Y Acad Sci* 2005;1064:37–49. [PubMed: 16394146]
3. Young K, Govindaraju V, Soher BJ, Maudsley AA. Automated spectral analysis I: formation of a priori information by spectral simulation. *Magn Reson Med* 1998;40(6):812–815. [PubMed: 9840824]

4. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*. Springer; New York: 2001.
5. Hand D. Classifier Technology and the Illusion of Progress. *Statistical Science* 2006;21(1):1–34. [PubMed: 17906740]
6. Taylor JE, Worsley KJ. Inference for magnitudes and delays of responses in the FIAC data using BRAINSTAT/FMRISTAT. *Hum Brain Mapp* 2006;27(5):434–441. [PubMed: 16568426]
7. Freeborough PA, Fox NC. MR Image Texture Analysis Applied to the Diagnosis and Tracking of Alzheimer's Disease. *IEEE Transactions on Medical Imaging* 1998;17(3):475–479. [PubMed: 9735911]
8. Rosen HJ, Gorno-Tempini ML, Goldman WP, Perry RJ, Schuff N, Weiner M, et al. Patterns Of Brain Atrophy In Frontotemporal Dementia and Semantic Dementia. *Neurology* 2002;58(2):198–208. [PubMed: 11805245]
9. Lerch JP, Evans AC. Cortical Thickness Analysis Examined Through Power Analysis and a Population Simulation. *Neuroimage* 2005;24(1):163–173. [PubMed: 15588607]
10. Bastos-Leite AJ, Scheltens P, Barkhof F. Pathological aging of the brain. *Top Magn Reson Imaging* 2004;15(6):369–89. [PubMed: 16041289]Review
11. Lerch JP, Pruessner JC, Zijdenbos A, Hampel H, Teipel SJ, Evans AC. Focal decline of cortical thickness in Alzheimer's disease identified by computational neuroanatomy. *Cereb Cortex* 2005;5(7):995–1001. [PubMed: 15537673]
12. Singh V, Chertkow H, Lerch JP, Evans AC, Dorr AE, Kabani NJ. Spatial patterns of cortical thinning in cognitive impairment and Alzheimer's disease. *Brain* 2006;129(Pt 11):2885–2893. [PubMed: 17008332]
13. Gold BT, Balota DA, Cortese MJ, Sergent-Marshall SD, Snyder AZ, Salat DH, Fischl B, Dale AM, Morris JC, Buckner RL. Differing neuropsychological and neuroanatomical correlates of abnormal reading in early-stage semantic dementia and dementia of the Alzheimer type. *Neuropsychologia* 2005;43(6):833–846. [PubMed: 15716156]
14. Liu AK, Marcus KJ, Fischl B, Grant PE, Young Poussaint T, Rivkin MJ, Davis P, Tarbell NJ, Yock TI. Changes in cerebral cortex of children treated for medulloblastoma. *Int J Radiat Oncol Biol Phys*. 2007In press
15. Calabrese M, Atzori M, Bernardi V, Morra A, Romualdi C, Rinaldi L, McAuliffe MJ, Barachino L, Perini P, Fischl B, Battistin L, Gallo P. Cortical atrophy is relevant in multiple sclerosis at clinical onset. *J Neurol*. 2007In press
16. Bearden CE, van Erp TG, Dutton RA, Tran H, Zimmermann L, Sun D, Geaga JA, Simon TJ, Glahn DC, Cannon T, Emanuel BS, Toga AW, Thompson P. Mapping Cortical Thickness in Children with 22q11.2 Deletions. *Cereb Cortex*. 20060 97.
17. Crutchfield JP, Young K. Inferring Statistical Complexity. *Physical Review Letters* 1989;63:105–107. [PubMed: 10040781]
18. Shalizi CR, Crutchfield JP. Computational Mechanics: Pattern and Prediction, Structure and Simplicity. *Journal of Statistical Physics* 2001;104:817–879.
19. Crutchfield, JP.; Young, K. Computation at the Onset of Chaos. In: Zurek, W., editor. *Entropy, Complexity, and Physics of Information, SFI Studies in the Sciences of Complexity, VIII*. Addison-Wesley, Reading; Massachusetts: 1990. p. 223-269.
20. Young K, Crutchfield JP. Fluctuation Spectroscopy. *Chaos, Solitons, and Fractals* 1993;4:5–36.
21. Young K, Chen Y, Kornak J, Matson GB, Schuff N. Summarizing complexity In High Dimensions. *Physical Review Letters* 2005;94:098701. [PubMed: 15784007]
22. Shannon CE. A mathematical theory of communication. *Bell Sys Tech Journal* 1948;27:379–423.
23. Cover, T.; Thomas, J. *Elements of Information Theory*. Wiley-Interscience; 2006.
24. Ripley, BD. *Statistical Inference for Spatial Processes*. Cambridge University Press; Cambridge, England: 1988.
25. Cressie, N. *Statistics for Spatial Data*. Wiley: Brisbane, Australia; 1993.
26. Marroquin JL, Vemuri BC, Botello S, Calderon F, Fernandez-Bouzas A. An accurate and efficient Bayesian method for automatic segmentation of brain MRI. *IEEE Trans Med Im* 2002;21(8):934–945.

27. Rousseeuw, A.; Struyf, M.; Hubert, R. Package: Functions for Clustering. 2004. Available at <http://cran.r-project.org/>
28. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magn Reson Imaging* 2004;22(1):81–91. [PubMed: 14972397]
29. Hopcroft, JE.; Ullman, JD. Introduction to Automata Theory, Languages, and Computation. Vol. 2. Addison Wesley; 2000.
30. Im K, Lee JM, Yoon U, Shin YW, Hong SB, Kim IY, Kwon JS, Kim SI. Fractal dimension in human cortical surface: multiple regression analysis with cortical thickness, sulcal depth, and folding area. *Hum Brain Mapp* 2006;27(12):994–1003. [PubMed: 16671080]
31. Feldman DP, Crutchfield JP. Structural Information in Two-Dimensional Patterns: Entropy Convergence and Excess Entropy. *Physical Review E* 2003;67:051104.
32. van Rossum, G.; Drake, FL. Python Reference Manual. PythonLabs; Virginia, USA: 2001. Available at <http://www.python.org/>
33. Jones, E.; Oliphant, T.; Peterson, P., et al. SciPy: Open Source Scientific Tools for Python. 2001. Available at <http://www.scipy.org/>
34. Moreira, W. RPy Package. 2004. Available at <http://rpy.sourceforge.net/>
35. R Development Core Team. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing. 2004. Available at <http://www.r-project.org/>
36. Dimitriadou, E.; Hornik, K.; Leisch, F.; Meyer, D.; Weingessel, A. R Package: e1071: Misc Functions of the Department of Statistics. 2004. Available at <http://cran.r-project.org/>
37. Marchini, JL. R Package: AnalyzeFMRI: Functions for Analysis of fMRI Datasets Stored in the ANALYZE Format. 2004. Available at <http://cran.r-project.org/>
38. Nielsen, O.; Ciceri, GP.; Ramachandran, P.; Orr, D.; Kaukic, M. PyPAR - Parallel Python, Efficient and Scalable Parallelism Using the Message Passing Interface (MPI). 2003. Available at <http://datamining.anu.edu.au/~ole/pypar/>
39. Mazziotta J, Toga A, et al. A Probabilistic Atlas and Reference System For the Human Brain: International Consortium for Brain Mapping (ICBM). *Phil Trans R Soc Lon, B Biol Sci* 2001;356:1293–1322.
40. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. Automated Anatomical Labeling of Activations In SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-subject Brain. *Neuroimage* 2002;15(1):273–289. [PubMed: 11771995]
41. Pluim JP, Maintz JB, Viergever MA. F-information measures in medical image registration. *IEEE Trans Med Imaging* 2004;23(12):1508–1516. [PubMed: 15575408]
42. Scott ML, Thacker NA. Robust tissue boundary detection for cerebral cortical thickness estimation. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv* 2005;8(2):878–885. [PubMed: 16686043]
43. Barta P, Miller MI, Qiu A. A stochastic model for studying the laminar structure of cortex from MRI. *IEEE Trans Med Imaging* 2005;24(6):728–742. [PubMed: 15957597]
44. Zaidi H, Ruest T, Schoenahl F, Montandon ML. Comparative assessment of statistical brain MR image segmentation algorithms and their impact on partial volume correction in PET. *Neuroimage* 2006;32(4):1591–607. [PubMed: 16828315]
45. Makris N, Hodge SM, Haselgrove C, Kennedy DN, Dale A, Fischl B, Rosen BR, Harris G, Caviness VS Jr, Schmahmann JD. Human cerebellum: surface-assisted cortical parcellation and volumetry with magnetic resonance imaging. *J Cogn Neurosci* 2003;5(4):584–99. [PubMed: 12803969]
46. Thirion B, Fugeras O. Feature characterization in fMRI data: the Information Bottleneck approach. *Image Anal* 2004;8(4):403–419.
47. Fischl B, Salat DH, vander Kouwe A, Makris N, Segonne F, Quinn BT, Dale AM. Sequence-independent Segmentation of Magnetic Resonance Images. *Neuroimage* 2004;23:S69–S84. [PubMed: 15501102]
48. Joshi S, Davis B, Jomier M, Gerig G. Unbiased Diffeomorphic Atlas Construction for Computational Anatomy. *Neuroimage* 2004;23:S151–S160. [PubMed: 15501084]

Figure 1.

Schematic representation of the methods for obtaining a joint histogram from multiple data sets coregistered over an index space. In the first step the values or “features” are mapped to a space that ignores the underlying index of the features. The features are then classified in the feature space and the class values (or class probabilities) are mapped back onto the indices. The resulting image is parsed and the histogram generated in the parsing step serves as an estimate of the joint and conditional pattern distributions. The entropy and statistical complexity are then estimated from the histogram.

Right Superior Temporal Gyrus (Cortical Thinning in 25 Subjects)

Figure 2.

Scatterplot matrix of H , SC , and EE , estimated for linear template length $L = 2$, for the right superior temporal gyrus (RSTG) region, for 50 subjects, 25 of which had the cortex in the RSTG artificially thinned while other brain regions were left unmodified. Those for which RSTG was thinned are represented by the open squares. The figure provides an illustration that H , SC , and EE , estimated at $L = 2$, automatically discovered to be optimal by the estimation procedure, do a good job at separating the group for which the cortex was thinned from the group for which there was no thinning.

Left Superior Temporal Gyrus (No Cortical Thinning)

Figure 3.

Scatterplot matrix of H , SC , and EE for linear template $L = 2$, estimated for the left superior temporal gyrus (LSTG) region, which was not modified in any subjects. Recall that half of the 50 subjects had the cortex artificially thinned in the RSTG but had no modification outside the RSTG, including the LSTG. The figure is separated into subjects for which the RSTG was artificially thinned and those which underwent no thinning in the RSTG. Subjects for which the cortex in the RSTG was thinned are represented by the open squares as in figure (2). The figure provides an illustration that there is no separation, as expected, in a region for which there was no thinning in contrast to the results for the RSTG shown in figure (2).

Table 1

Classification analyses using *H*, *SC*, and *EE* from whole brain (WB), temporal lobe (TL), superior temporal gyrus (STG), and right and left superior temporal gyrus (RSTG, LSTG)

	MANOVA <i>p</i>	LDA <i>p</i>	SENSITIVITY	SPECIFICITY
WB	0.2	0.03	0.59	0.52
TL	0.002	0.0002	0.73	0.63
STG	6e-06	1.7e-07	0.80	0.76
LSTG	0.2	0.03	0.60	0.55
RSTG	4e-09	5.7e-11	0.91	0.85

p-values for Wilks lambda from a MANOVA and Welch two sample t-test on the first component from a linear discriminant analysis (LDA), and sensitivity and specificity using 10 times 10-fold cross validation on the LDA results. Note that, using LDA and the complexity estimates in the RSTG, a sensitivity of 0.91 was achieved as compared to 0.93 for the best technique described in [11] for directly estimating cortical thickness in the RSTG.