

Statistical Inference Models for Image Datasets with Systematic Variations

Won Hwa Kim^{†¶} Barbara B. Bendlin^{*¶} Moo K. Chung[§] Sterling C. Johnson^{*¶} Vikas Singh^{§†¶}

[†]Dept. of Computer Sciences, University of Wisconsin, Madison, WI

[§]Dept. of Biostatistics & Med. Informatics, University of Wisconsin, Madison, WI

[¶]Wisconsin Alzheimer's Disease Research Center, University of Wisconsin, Madison, WI

^{*}GRECC, William S. Middleton VA Hospital, Madison, WI

Abstract

Statistical analysis of longitudinal or cross sectional brain imaging data to identify effects of neurodegenerative diseases is a fundamental task in various studies in neuroscience. However, when there are systematic variations in the images due to parameter changes such as changes in the scanner protocol, hardware changes, or when combining data from multi-site studies, the statistical analysis becomes problematic. Motivated by this scenario, the goal of this paper is to develop a unified statistical solution to the problem of systematic variations in statistical image analysis. Based in part on recent literature in harmonic analysis on diffusion maps, we propose an algorithm which compares operators that are resilient to the systematic variations. These operators are derived from the empirical measurements of the image data and provide an efficient surrogate to capturing the actual changes across images. We also establish a connection between our method to the design of wavelets in non-Euclidean space. To evaluate the proposed ideas, we present various experimental results on detecting changes in simulations as well as show how the method offers improved statistical power in the analysis of real longitudinal PIB-PET imaging data acquired from participants at risk for Alzheimer's disease (AD).

1. Introduction

Statistical analysis of a cohort of brain imaging scans to assess the long term effects of trauma/stress and identify genetic, demographic and lifestyle factors for neurodegenerative diseases is a cornerstone of current research in neuroscience. Typically, the population will consist of two clinically disparate groups/classes: say, diseased and healthy (cross-sectional) or a set of subjects imaged several years apart (longitudinal). Once all images are 'registered' to a common template space, the statistical analysis can proceed in a number of ways. For instance, at each voxel one may

perform a hypothesis test (e.g., Student's t -test) to ask if the distribution of intensities at *that* voxel across the two distinct groups are the same [10]. If there is sufficient evidence to reject the null hypothesis, we can conclude with some confidence (0.05 α level) that the voxel is relevant for the disease. By repeating this procedure across all voxels, we can obtain a heat map of p -values to identify potential regions affected by the disease.

There are two basic but important issues we should emphasize here. First, our ability to conclude that (at a specific voxel) the observed empirical intensity distributions are different across groups depends on the sample size and how distinct the distributions are (i.e., the effect size). Second, this analysis assumes that the *absolute* image intensity measurements are meaningful. In other words, we assume that the only differences between the groups is due to the effect of the clinical phenomena under study (i.e., age, disease and so on), and *not* other global systematic variations coming from modifications in acquisition parameters. Generally, in small to medium sized studies where the data is acquired at a single site (with the same scanner), this is not a problem. But as scientific studies investigate more subtle scientific questions where the group differences are *weaker*, we need larger sample sizes — logistic constraints necessitate multi-site studies. Changes in the hardware and pulse sequences (and many other factors) across sites introduce systematic variations in the dataset. In fact, even in small studies, a hardware upgrade (between baseline and follow-up acquisitions) may be a nuisance for analysis, requiring ad-hoc normalization which may affect statistical power of detecting true group effects. When the effect sizes are poor, performing inference on the data without appropriate adjustments could affect the success or failure of the scientific hypothesis under investigation.

The above problem is common across various imaging modalities in medical imaging. For instance, in neuroimaging uses of positron emission tomography (PET), a nuclear imaging modality (where an injected radiotracer binds to

specific pathologies), image measurements vary considerably, even for the same subject, due to a variety of reasons. So, before any statistical analysis can be performed, these images must be “normalized”. Possible approaches include global normalization (mean intensity) or regional scaling (by a reference region). This process converts the intensities into a physiological range of interpretable values. But if the global average or the mean intensity of the reference region used is *not* independent of the condition being studied, the analysis will invariably suffer. In these cases, incorrect normalization can lead to an inability to identify real group differences, or worse, one may obtain paradoxical or “opposite” findings. In various other imaging modalities, a normalization strategy may not even be viable. For example, if the systematic variations are the result of changes in the acquisition parameters at different sites, one must analyze the smaller datasets separately. The goal of this paper is to develop a unified statistical solution to this problem.

A high level description of the strategy. Let f denote an unknown function. Let α and β denote two parameters such that they modify the form of the function $f(\cdot)$ yielding f_α and f_β . Now, consider that we are only given access to measurements of f_α and f_β . It is clearly not possible to verify whether they were both derived from the same latent function, f unless we also know the relationship between the transformations of f induced by α and β (if the respective inverse transformations are unique). Assume that an oracle provides us an operator \mathcal{T} (to be described in detail) with the interesting property that it is *invariant* to the parameter space \mathcal{P} from which α and β are drawn. That is, if we construct a pair of operators from the empirical measurements of f_α and f_β , the operators will be the same: $\mathcal{T}_{f_\alpha} \equiv \mathcal{T}_{f_\beta}$ if they share the same latent function, f .

Next, consider a slightly more complicated setup. The latent function f has now been modified to f' . We are now provided with the measurements, f_α and f'_β , i.e., both the parameter and the function change. Since the operator \mathcal{T} only offers invariances to the parameter space \mathcal{P} (and assumes that the latent function is the same), in this case, the operators \mathcal{T}_{f_α} and $\mathcal{T}_{f'_\beta}$ *cannot* be compared. Nonetheless, we can see that the operators provide a mapping to two *different* spaces, say S_{f_α} and $S_{f'_\beta}$, since f and f' are distinct. Interestingly, because of the invariance to \mathcal{P} , if we now plug in a *known* function (such as an impulse function) at all locations in the original space into the two operators, we will obtain its transformed representations in S_{f_α} and $S_{f'_\beta}$. Once these transformed forms of the impulse functions are mapped from $S_{f'_\beta}$ to S_{f_α} , we can calculate the distance. If the distance is near zero, then $f \simeq f'$; otherwise, it characterizes the discrepancy between f and f' since the operators are, by design, invariant to \mathcal{P} .

The main **contribution** of this paper is to formalize this idea for immunity to systematic variations in statistical

analysis of imaging data, based on a new method in the harmonic analysis literature by Coifman and Hirn [6]. In particular, we **a)** Derive the operators \mathcal{T} using the recent work in Diffusion Maps [7, 6] and show how the corresponding invariance allows performing statistical analysis of systematically varying images, i.e., f_α and f_β for $\alpha, \beta \in \mathcal{P}$. Note that it may not otherwise be possible to even compare f_α and f_β ; **b)** Describe how the lower dimensional mapping obtained by the operators relate to a Wavelet transform in non-Euclidean spaces; **c)** Provide experimental evidence in that the method facilitates statistical analysis of Pittsburgh compound B PET (PIB-PET) [20] images and offers improvements over standard normalization methods used in the community.

1.1. Related Work

There are several broadly related ideas in vision and medical imaging that can serve as a reasonable starting point for comparing functions that cannot otherwise be compared [8]. The most natural choice is a statistical measure that is, by construction, invariant to image intensities: Mutual Information (MI). Mutual information has been extensively used in both computer vision (e.g., stereo [12, 16, 13]) and in medical imaging (e.g., non-linear registration [32, 22, 26, 33]) and offers precisely the type of invariance we desire. While MI is a good loss function to optimize when searching for a non-linear transformation or disparity map, once such a transformation has been found and the images have been aligned, MI does not make the statistical analysis any easier. For instance, consider a set of ten participants whose images were acquired twice, a few years apart, and the intensities in the second acquisition are systematically different (e.g., affine scaling). While MI *can* characterize the joint entropy of a pair of intensities, it *cannot* be used to quantify the voxel-wise change from one time point to the other.

An alternative to the MI approach is based on dictionary learning/patch regression inspired idea called *image synthesis* [14, 28, 25]. Broadly, one may use image synthesis to synthetically generate the image that has been corrupted, assuming a large set of training examples is available. While this approach is suitable for addressing missing data, applying it in the above longitudinal setting will entail generating the *entire* set of images at the second time point. The learning task will broadly correspond to inferring the parameters of a generative model that explains *temporal change* across the population, given only the baseline acquisition — clearly difficult regardless of how well characterized the training dataset is. Given these issues, to our knowledge, there is no universally applicable solution offering the same capabilities as the algorithm we propose here. In situations where the structural variations in the intensities are related by a simple transformation, one may normalize the entire image by a suitable normalization constant. In medical

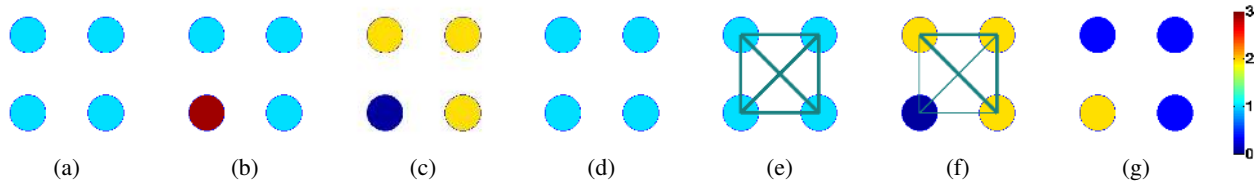


Figure 1: Comparing two different functions on four data points. a) $f(p) = f_\alpha(p) = (1, 1, 1, 1)^T$, b) $f'(p) = (1, 1, 3, 1)^T$, c) $f'_\beta(p) = 3 - f'(p) = (2, 2, 0, 2)^T$, d) $|f_\alpha - f'_\beta|$, e), f) graphs from f_α and f'_β respectively (edge thickness denotes to edge weight), g) WKD using structure from (e) and (f). The true change between (a) and (b) is $(0, 0, 2, 0)$, but a simple subtraction in (d) is inaccurate. The proposed algorithm can capture the true change in (g).

imaging, this is often difficult because it must be derived from a region not affected by disease, age, or the clinical phenomena under study. If this is sub-optimal, it can affect the statistical analysis in unexpected ways. Later, we will show experiments using the above approach as a baseline and demonstrate how the new method offers improvements.

2. A Multi-resolutional Perspective

Our framework considers the case where f_α and f_β correspond to images. Here we assume that the two images are spatially registered so that the *only* variations in the measurements comes from the parameters, $\alpha, \beta \in \mathcal{P}$. Although we cannot compare f_α and f_β due to parameter differences, the latent function is the same, so the overall behavior of the functions are similar modulo the variation introduced by the parameters drawn from \mathcal{P} . Formally, we assume that the relationship of measurements at *each* grid point to measurements at *other* grid points in the same image is preserved and is independent of changes in the parameter space, \mathcal{P} . In other words, if we place a unit mass of energy at position p in the first image, then the overall pattern of dissipation to its surroundings (e.g., governed by the form of the function) will be similar to that observed in the second image. Now, imagine that f_α changes to f'_α , which alters the relationship between the sample points in subtle ways. When we place a unit energy at location (or grid point) p , the propagation of the energy will now show *different* patterns for the two image-derived operators: capturing the difference between those patterns is an excellent surrogate for detecting the difference between the two original functions f and f' .

In the previous section, we proposed that this procedure should be carried out by operators \mathcal{T}_{f_α} and $\mathcal{T}_{f'_\beta}$ derived from the measurements of respective functions. We know that these operators define a mapping to lower dimensional spaces S_f and $S_{f'}$ [2]. The unit energy at position p can be simulated by an impulse function δ_p . Now, when \mathcal{T}_{f_α} and $\mathcal{T}_{f'_\beta}$ are applied to the same δ_p , the energy propagation will be characterized in the spaces S_f and $S_{f'}$. Interestingly, the reader will see that this process is exactly like the construction of a mother wavelet function [23, 30], which involves applying a Wavelet transformation operator on a delta function to characterize dissipation of a unit energy in a lower dimensional space. Traditionally, a mother wavelet function $\psi_{s,p}$ corresponds to scale s and location p — while the for-

mer defines the dilation, the latter corresponds to individual sample grid points in the image. This discussion immediately suggests the possibility of using the Wavelet transform in place of the operator, \mathcal{T} .

One important issue here is that wavelets are classically defined in \mathbb{R}^n , where the space is represented as a grid graph with *equally weighted* connections to the neighbors. Instead, in this application, we want the energy to spread *unevenly* in all possible directions, strongly modulated by the strength of the connections between the data points. This corresponds to a *non-Euclidean space* where the design of Wavelets is poorly investigated. Shortly, we will describe how the wavelet transform on graphs [11, 7] provides a solution to this problem by building Wavelet basis in such arbitrary structured spaces.

The Spectral Graph Wavelet transform (SGWT) [11]. The construction of a wavelet transform is problematic in a non-Euclidean space because of the two main properties of the Wavelet bases: translation and scaling. For instance, because of the arbitrary connections between the grid-points, it is difficult to imagine what the “shift” or “dilation” of a function means. But instead of defining the mother wavelet in the original domain, wavelets can be simply defined as band-pass filters in the frequency domain. SGWT defines a Wavelet basis based on spectral graph theory to first transform the data into the frequency domain. Then, operations analogous to scaling can be accomplished by band-pass filtering and then transforming the data back — this one shot backward/forward process gives a Wavelet basis for graphs.

Consider a graph $\mathcal{G} = \{V, E, \omega\}$ defined by a vertex set V , an edge set E and corresponding edge weight ω . A convenient representation of the graph is a $N \times N$ adjacency matrix A where N is the number of vertices and each element a_{ij} in A denotes the edge between the i th and j th vertices by the weight ω . A diagonal $N \times N$ matrix D , known as degree matrix, characterizes the degree of each vertex, as the i th diagonal element is defined as $\sum_{(i,j) \in E} a_{ij}$. From these two matrices, a graph Laplacian L is derived as $L = D - A$ which is a self-adjoint operator and upon decomposition provides a full set of orthonormal bases. SGWT uses these orthonormal bases to define the graph Fourier transformation which in turn provides the graph Wavelet transformation.

Based on the given λ and χ , SGWT is determined by a

wavelet kernel function $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, which behaves as a band-pass filter. Then, the SGWT operator \mathcal{T}_g on a function $h(p)$ is defined as

$$(\mathcal{T}_g h)(p) = \sum_{l=0}^{N-1} g(\lambda_l) \beta_l \chi_l(p) \quad (1)$$

where $\beta_l = \langle h, \chi_l \rangle$. Note that there is no notion of scaling in (1). But we know that the Fourier transform has a scaling property which can be incorporated in the kernel function, $g(\cdot)$. Then, the corresponding operation (but now with scaling property) is given as

$$W_h(s, p) = (\mathcal{T}_g^s h)(p) = \sum_{l=0}^{N-1} g(s\lambda_l) \beta_l \chi_l(p). \quad (2)$$

which results in the wavelet coefficient $W_h(s, p)$. The actual mother wavelet $\psi_{s,p}$ at scale s centered at p is realized by a delta function δ_p as

$$\psi_{s,p}(q) = \mathcal{T}_g^s \delta_p = \sum_{l=0}^{N-1} g(s\lambda_l) \chi_l(p) \chi_l(q) \quad (3)$$

and this satisfies the conventional transform $W_h(s, p) = \langle \psi_{s,p}, h \rangle$. This wavelet function is considered as a kernel function $\psi_s(p, q)$ between vertices p and q , serving as a key ingredient in defining the notion of “change” between f_α and f'_β independent of the parameters $\alpha, \beta \in \mathcal{P}$

3. Change Detection in Non-Euclidean Space

3.1. Wavelet Map and Wavelet Kernel Distance

Defining a kernel function in a square integrable measure space (X, μ) enables one to measure local similarities within X at small scales [6]. In our case, we define a mother wavelet function as such a kernel function using an operator \mathcal{T}^s . The operator \mathcal{T}^s is constructed using empirical measurements of function f_α which are given. A wavelet function $\psi_{s,p}(q)$ can be viewed as if it were a kernel function written as $\psi_s(p, q)$, defining a relationship between vertex p and q [3]. Following [6], we can define a kernel function $\psi_s(p, q)$. For our application, this yields Wavelet Kernel Distance (WKD) $d_s(p, q)$ at scale s , a measure between two points p and q defined as ℓ^2 -norm of the wavelet density difference over the space X as

$$d_s(p, q)^2 = \|\psi_s(p, \cdot) - \psi_s(q, \cdot)\|_2^2 \quad (4)$$

$$= \int_X (\psi_s(p, r) - \psi_s(q, r))^2 \mu(r) \quad (5)$$

In the graph setting, using the SGWT operator \mathcal{T}_g^s defined by a set of a set of eigenvalues and eigenfunctions as in Section 2, observe that (4) can be rewritten using a wavelet kernel function $g(\cdot)$ in the spectral domain as

$$d_s(p, q)^2 = \sum_{l=0}^{N-1} g(s\lambda_l)^2 (\chi_l(p) - \chi_l(q))^2 \quad (6)$$

The expression in (6) lies at the heart of the idea developed in this paper. It can be interpreted as if we were comparing the effect of the same wavelet function dissipating from different locations p and q to their neighbors by the wavelet kernel function $g(\cdot)$, thereby measuring the effect of the propagation. Further, we can also define a mapping of δ_p at each grid-point to a lower dimensional Euclidean space spanned by χ defined as the wavelet map $\gamma : X \rightarrow \ell^2$ at scale s as

$$\gamma_s(p) = (g(s\lambda_l) \chi_l(p))_{l=0,1,\dots,N-1} \quad (7)$$

characterizing the local relationship of the graph with the wavelet kernel function $g(\cdot)$. Note that when $g_s(\lambda_l) = \lambda_l^s$, the wavelet map exactly becomes diffusion map proposed earlier in [7].

A toy example is shown in Fig. 1: the objective here is to compare two different functions f_α and f'_β defined on four data points, and find the true difference between them. Given latent functions $f = (1, 1, 1, 1)^T$ and $f' = (1, 1, 3, 1)^T$, the true difference (i.e., $|f - f'|$) here is $(0, 0, 2, 0)$. Given the latent functions, f_α remains the same as f while f'_β is defined to be $f'_\beta(p) = 3 - f'(p)$. Clearly, a direct comparison of f_α and f'_β (i.e., $|f_\alpha - f'_\beta|$), as illustrated in Fig. 1 d), fails to detect the true difference. On the other hand, computing WKD from graphs constructed using f_α and f'_β at each data point yields the true difference as shown in Fig. 1 g).

We can now formally establish the relationship between wavelet map, WKD, and the construction of Wavelets using the following two results.

Proposition 1. *The squared WKD d_s defined between two vertices p and q on the same graph is equivalent to the ℓ^2 -norm of the difference between the respective wavelet maps of vertices p and q .*

Proof. Taking the ℓ^2 -norm of the difference over wavelet map on vertices p and q yields,

$$\begin{aligned} d_s(p, q)^2 &= \|\gamma_s(p) - \gamma_s(q)\|_2^2 = \sum_{l=0}^{N-1} (g(s\lambda_l) \chi_l(p) - g(s\lambda_l) \chi_l(q))^2 \\ &= \sum_{l=0}^{N-1} g(s\lambda_l)^2 (\chi_l(p) - \chi_l(q))^2. \quad \square \end{aligned}$$

From Proposition 1, we can see that WKD defines a Euclidean distance of the wavelet maps between vertices p and q in the space formed by χ . But is there a relationship between Wavelet maps and an actual wavelet function?

Proposition 2. *Let $\mathcal{V} = [\chi_0 \ \chi_1 \ \dots \ \chi_{N-1}]$ denote a matrix where χ_i corresponds to columns. The projection of a wavelet map $\gamma_s(p)$ at vertex p to the row space of \mathcal{V} precisely constructs a mother wavelet function $\psi_{s,p}(q)$.*

Proof. Given $\chi(q)$, the q th row of \mathcal{V} , taking inner product of the wavelet maps $\gamma_s(p)$ and $\chi(q)$ becomes

$$\langle \gamma_s(p), \chi(q) \rangle = \sum_{l=0}^{N-1} g(s\lambda_l) \chi_l(p) \chi_l(q) = \psi_{s,p}(q)$$

which defines a wavelet function at q centered at p exactly in the form given in (3). \square

Summary. We see that Proposition 2 establishes the connection between the construction of Wavelet functions from Section 2 and the wavelet map. It shows that a Wavelet function can be constructed from the wavelet map at each vertex. Further, this result ties the wavelet map to kernel signatures on graphs, variants of which have been used for graph matching and surface segmentation (but using diffusion [9, 31]). When the wavelet map of p is projected to the p th row of \mathcal{V} , we get a wave-type kernel descriptor in [1, 17]. Separately, when $g_s(\lambda_l) = \exp(-s\lambda_l)$, we obtain the heat kernel signature in [4].

3.2. Generalization of Wavelet Kernel Distance

So far, we have shown how two different vertices on the same image/graph can be compared using a Wavelet operator \mathcal{T}_{f_α} that has been derived from empirical measurements of a function f_α . But our main interest in facilitating statistical analysis of longitudinal systematically varying data is in comparing the same measurement location across the two images. We now derive such a generalization.

Consider two individual graphs I and J , constructed using functions (or images) f_α and f'_β , where the number of vertices in each is N . We assume that the vertices are spatially registered and that we are operating on a square integrable space X . On these graphs, WKD between a vertex p^I from I and a vertex q^J on J is defined as

$$d_s(p^I, q^J)^2 = \|\psi_s^I(p, \cdot) - \psi_s^J(q, \cdot)\|_2^2 \quad (8)$$

$$= \int_X (\psi_s^I(p, r) - \psi_s^J(q, r))^2 \mu(r) \quad (9)$$

using wavelet kernel functions ψ_s^I and ψ_s^J .

Our basic recipe is to construct two operators \mathcal{T}_{f_α} and $\mathcal{T}_{f'_\beta}$, and obtain two sets of orthogonal bases χ^I and χ^J from each operators to compare the vertex-wise differences. Note that while the expansion of (8) does not simplify as in (6) since the eigenvectors χ^I and χ^J are no longer orthogonal to each other, it nonetheless reduces to a meaningful expression defining a mapping between the lower dimensional spaces defined by the two operators as described by the following result.

Proposition 3. *Let λ^I, λ^J and χ^I and χ^J denote the eigenvalues and eigenvectors from graphs of I and J respectively.*

Then, the WKD $d_s(p^I, q^J)$ can be written as,

$$d_s(p^I, q^J)^2 = \sum_{l_1=0}^{N-1} g(s\lambda_{l_1}^I)^2 \chi_{l_1}^I(p)^2 + \sum_{l_2=0}^{N-1} g(s\lambda_{l_2}^J)^2 \chi_{l_2}^J(q)^2 \quad (10)$$

$$- 2 \sum_{l_1, l_2=0}^{N-1} g(s\lambda_{l_1}^I) \chi_{l_1}^I(p) g(s\lambda_{l_2}^J) \chi_{l_2}^J(q) \langle \chi_{l_1}^I, \chi_{l_2}^J \rangle$$

The proof of Proposition 3 is given in the extended version. It is instructive to tease apart the various terms in (10) to understand their behavior. The first two terms in (10) form the WKD on a single graph whereas the last term compensates for the discrepancy caused by the variations of the inherited spaces once the first space has been mapped to the other. By inspection, we see that this generalizes Proposition 1. When I and J are the same, we can verify,

Proposition 4. *When I and J are equal, then (10) reduces to (6).*

Proof. Since I and J are the same graph, they share the eigenvalues λ_l and eigenvectors χ_l , therefore

$$d_s(p^I, q^J)^2 = \sum_{l_1=0}^{N-1} g(s\lambda_{l_1}^I)^2 \chi_{l_1}^I(p)^2 + \sum_{l_2=0}^{N-1} g(s\lambda_{l_2}^J)^2 \chi_{l_2}^J(q)^2$$

$$- 2 \sum_{l_1, l_2=0}^{N-1} g(s\lambda_{l_1}^I) \chi_{l_1}^I(p) g(s\lambda_{l_2}^J) \chi_{l_2}^J(q) \langle \chi_{l_1}^I, \chi_{l_2}^J \rangle$$

$$= \sum_{l=0}^{N-1} g(s\lambda_l)^2 (\chi_l(p)^2 - \chi_l(q)^2)$$

with $\langle \chi_{l_1}^I, \chi_{l_2}^J \rangle = 0$ when $l_1 \neq l_2$. \square

4. Experimental Result

We demonstrate three sets of experimental results: two of these correspond to a situation where the ground truth variations are *known* whereas the third one is focused on a real statistical analysis problem on brain imaging data (real longitudinal PIB images). The first experiment evaluates whether the proposed method can detect actual changes across two different images where there are substantial systematic variations. In the second experiment, we carry out a statistical group analysis on synthetically generated PIB images, where we expect to detect true group differences from a model of the first and the second groups, in the presence of systematic variations. Finally, we run the proposed algorithm on a real longitudinal PIB-PET image dataset. Here, the images *are* normalized in a certain sense, however, because of the characteristics of the imaging modality we still expect systematic variations (depending on the accuracy of the normalization) which decreases the statistical power. Our goal here is to detect those regions in the brain that show high correlation between (a) the PIB changes over time and (b) known risk factors for Alzheimer's disease. Since the ground truth here is unavailable, we expect regions identified by our method to be consistent with those reported in the literature.

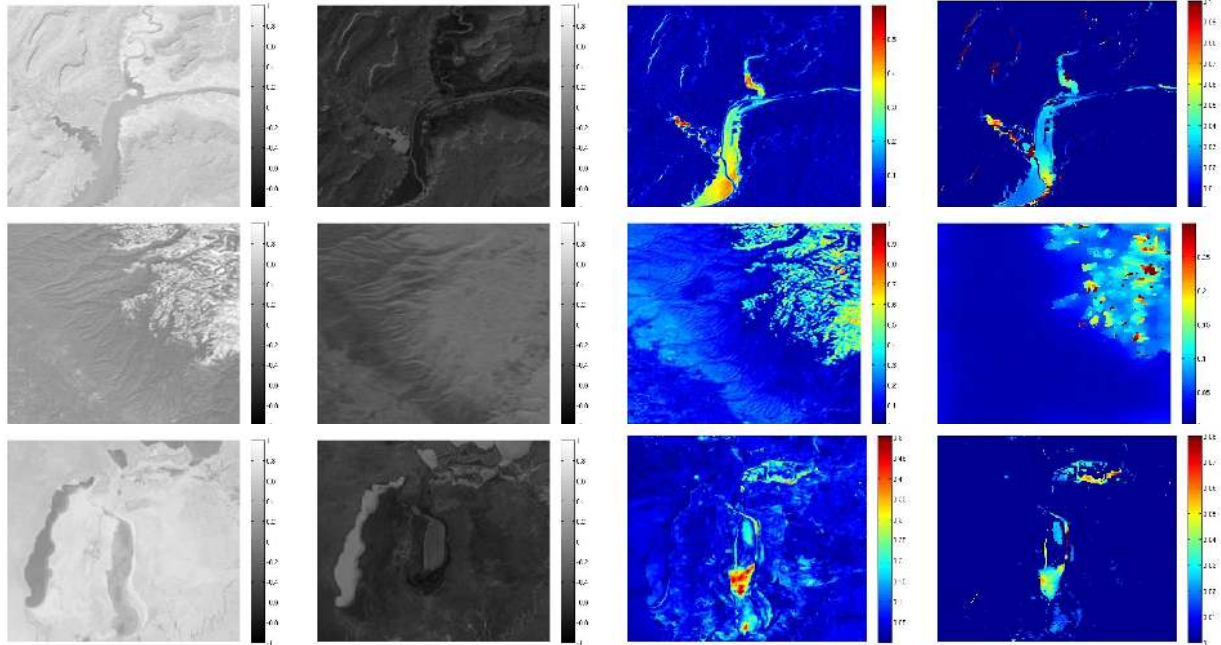


Figure 2: Results from NASA Earth Observatory images. We detect the changes between two images (in different scales) of Lake Powell (first row), Sierra Nevada (second row) and Aral Sea (third row). First column: images taken in 2013, Second column: images taken in 2014 (inverted), Third column: ground truth, Fourth column: changes identified using WKD.

4.1. Simulation on NASA Satellite Images

We obtained real satellite images from NASA Earth Observatory (<http://earthobservatory.nasa.gov>), which shows changes (due to various factors) over time at various locations worldwide. For each scene, we have two longitudinally acquired images which should reveal how the region has changed over time. To simulate ‘systematic variations’, we invert one of the images by multiplying all intensities by -1 , providing f_α and f'_β . Notice that direct comparison of these images yields nonsensical results. One can use Mutual information to derive a joint entropy of each pair of intensity values. Unfortunately, this scheme does not directly yield a Δt -image showing change over time. The key here is to notice that instead of comparing pixel intensities, we are detecting changes of local structures at each pixel between the two images in a lower dimensional Euclidean space, therefore we *are* able to identify the high-level differences meaningfully. Representative results are demonstrated in Fig. 2.

4.2. Group Analysis of Synthetic PIB images

We now present results of statistical analysis on a population of synthetically generated 2-D Pittsburgh Compound B (PIB) image data. The experiment design is as follows. We assume we have two groups: diseased and healthy (controls). We simulate brain images of 20 diseased and 20 control subjects, using a template 2-D PIB image with size of 79×95 . We assume that each subject was imaged longitudinally providing a t_0 (baseline) and t_1 (follow-up) image.

At t_0 , the images Y_{t_0} in both (diseased and control) groups are modeled as a random field with mean μ_{control} with added Gaussian noise $N(0, 0.1)$ as

$$Y_{t_0} = \mu_{\text{control}} + N(0, 0.1) \quad (11)$$

where μ_{control} is given by the template PIB image slice shown in Fig 3 (a). At t_1 , we consider two types of changes: the first is an increase of PIB values by 20% in certain regions of the brain in the diseased group characterized by μ_{disease} , and the other is systematic variation simulated as an arbitrary affine transformation with scale $s \in [1, 2]$ and translation $a \in [0, 1]$ applied to the image intensities.

$$Y_{t_1} = \begin{cases} s\mu_{\text{control}} + a + N(0, 0.1) & \text{if normal} \\ s\mu_{\text{disease}} + a + N(0, 0.1) & \text{if diseased} \end{cases}$$

In this scenario, we would like to detect the changes $\Delta Y = Y_{t_1} - Y_{t_0}$ from the two time points across the two groups by comparing the distribution of ΔY across groups. In the standard procedure, performing a statistical hypothesis test at each pixel (a total of 7505 tests) yields a p -value at each pixel, that tells us whether the distribution of the ΔY are the same. Applying Bonferroni correction at 0.05 removes false positives and identifies the regions with significant changes between the two groups. This process works well when $s = 0$ and $a = 0$, however, systematic variations may reduce or bias the effect sizes and diminish the statistical power. Using our method, we expect to detect the group differences even in the presence of systematic variation.

The resultant p -value maps from this simulation is displayed in Fig. 3 (b), (c) and (d) at the same scale ($-\log_{10}$

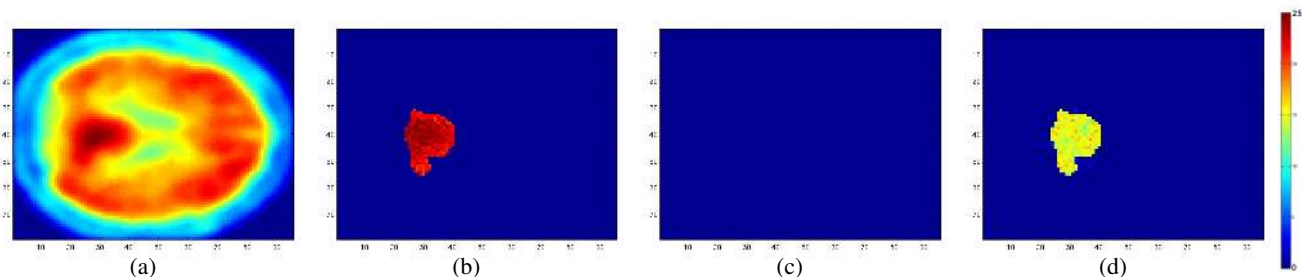


Figure 3: Result from a group analysis on diseased vs. normal groups using synthetic PIB images. a) a template PIB image used for the mean μ , b) p -value map in $-\log_{10}$ scale from the group analysis using images without the systematic variation (serving as the ground truth), c) p -value map in $-\log_{10}$ scale from the group analysis using images with systematic variation, d) p -value map in $-\log_{10}$ scale from the group analysis using WKD on images with systematic variation. We can see that using WKD, we can detect group differences even when there is a systematic variations in the images.

scale), which shows three cases of this experiment: using the standard hypothesis testing procedure on (i) the given data without systematic variations (i.e., ground truth), (ii) with systematic variations and (iii) WKD for the data with systematic variations. As seen in Fig. 3 (b), there is a strong signal showing group differences between the two groups (diseased and controls), easily identified using standard hypothesis testing. This serves as the ground truth. In contrast, when there are systematic variations in the data, the traditional approach fails to detect the true differential signal as shown in Fig. 3 (c). We computed WKD at each pixel of the images with systematic variations instead of computing ΔY directly, and then applied hypothesis testing on WKD. This process successfully detects the region as shown in Fig. 3 (d) showing excellent consistency with the actual changes between t_0 and t_1 . Therefore, in this sanity check experiment, our method correctly picks up the true variations and makes the downstream statistical analysis more sensitive even when systematic variations exist.

4.3. Analysis of Longitudinal PIB Changes

In this section, we demonstrate results from a longitudinal PIB-PET image analysis, where we use the ratio of total τ protein and amyloid- β -142 ($A\beta(1-42)$) as a predictor for the increase in voxel-wise PIB values at two different time points. PIB values are used as a measure of brain amyloid deposition, a core pathological feature of Alzheimer’s disease (AD), and it is known that such increase is closely correlated with AD. The $A\beta(1-42)$ interacts with the signaling pathways to control the phosphorylation of τ protein [21, 15] and their ratio is widely used as a sensitive feature of AD pathology.

Dataset: The dataset of 84 participants used here includes subjects that are otherwise healthy but may have potential risk factors for AD. The cohort is comprised of 26 males and 58 females, and the mean age is 67.4. The PIB images are a 3-D volume spatially registered to the Montreal Neurological Institute (MNI) space, then blank boundaries were cropped to obtain images of size $79 \times 95 \times 68$. The image intensities represent standard uptake value

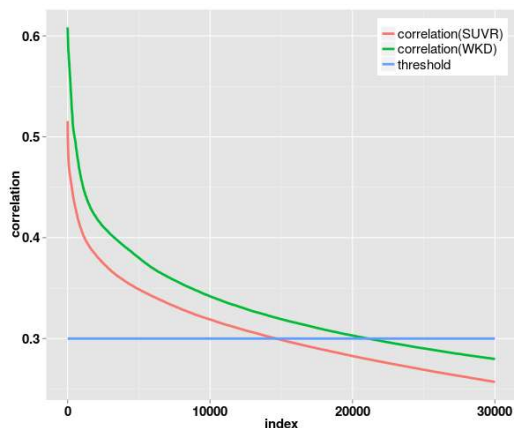


Figure 4: Plot of sorted correlation (descending) with respect to the number of voxels. The correlation using WKD (green) and SUVR images (red) show that WKD shows stronger correlation and larger number of vertices above the threshold (blue).

(SUV), which is the ratio of the tissue radioactivity concentration and injection divided by the body weight. These values are scaled with the intensity from a reference region (i.e., cerebellum), generating standard uptake value ratio (SUVR) images. The PIB intensities, by nature, *only increase* when affected by a disease factor. However, when the SUVR images (normalized using a reference region) between two time points are compared, various brain regions show decrease in the PIB values. This suggests that there *are* systematic variations in the two images that have *not* been account for by the normalization process.

Experimental setup: For the graph representation of each volume image, we used a grid graph with six neighbors for each voxel in 3-D space. The connection between voxels were defined by $\exp(-\|I(p) - I(q)\|^2/\sigma^2)$ where $I(p)$ is the PIB intensity at voxel p and $\sigma = 0.1$. The graph Laplacian for each subject had dimension of 510340×510340 , which was too big for standard solvers, we therefore used a Jacobi-Davison conjugate gradient method [24] to compute the first fifty eigenvalue/eigenvector pairs of the matrix. For the wavelet kernel function g , we used the cubic spline func-

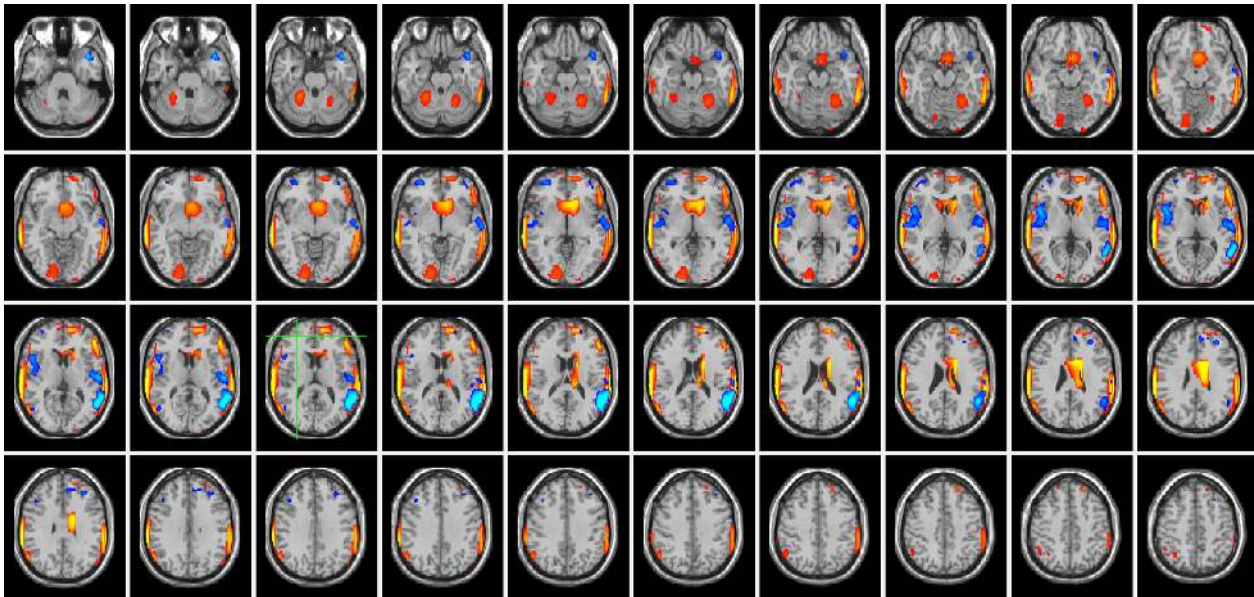


Figure 5: Montage of axial view of the correlation between the PIB changes and the ratio of total τ -protein and $A\beta(1-42)$ on a template T1-weighted brain image. The red-yellow intensities indicate correlation using WKD, and the blue-light blue intensities indicates correlation using SUVR images in the range of $[0.3 \ 0.5]$.

tion provided in SGWT [11].

Result: A high positive correlation between the PIB changes and the ratio between total τ protein and $A\beta(1-42)$ indicates that the increase of the PIB values are highly related to the increase of the ratio. When compared to the result using SUVR images, the correlation from WKD is stronger, and we also find larger regions of the brain. To quantitatively compare them, among the total of 510340 voxels, WKD identifies 21101 voxels (4.13%) with correlations above 0.3 — a common threshold for moderate correlation. On the other hand, using SUVR images, we find only 14655 voxels (2.87%) above 0.3. These correlations are sorted and shown in Fig. 4, indicating that WKD is more sensitive than the differences found via SUVR images.

Fig. 5 shows the resultant correlation overlaid on a T1-weighted template, where the correlations using WKD and SUVR images are shown in red-yellow and blue-light blue maps in the same range respectively. The result shows that both our analysis and the one performed on SUVR images agree on moderate correlations in lateral temporal lobe regions, which are well-known to be affected by AD [5, 29, 18, 19] — but our algorithm shows higher correlation and larger regions. Interestingly, WKD framework also picks up the bilateral cerebellum regions which is known to show loss of volume with dementia [27]. Note that this region is very close to regions that are used as the ‘reference’ for the SUVR normalization — therefore will not be identified in the standard analysis even if affected by disease.

Remark. There are some potential limitations of the method from the neuroscience point of view. For instance, one issue is that the analysis may miss out on some regions that are found by the standard analysis. In these situations, it

is difficult to assess whether this is an artifact of our method or a consequence of the normalization process in the standard analysis. We believe that a conservative option is to use our proposed algorithm as the first stage of analysis, which can be followed up by more specific region of interest based approaches common in neuroimaging.

5. Conclusion

This paper provides a solution to a problem where statistical analysis of imaging data in brain imaging studies is problematic due to systematic variations caused due to a variety of factors. Motivated from recent literature in harmonic analysis, we propose to compare operators as a means of detecting changes across images, when the absolute measurements cannot be compared on their own. These operators are derived from empirical measurements of images and provide invariance to the systematic variations. Using our framework, we showed experiments on synthetic as well as real datasets, demonstrating that the algorithm works well in a regime where few alternatives are currently available. In particular, in an interesting application to brain imaging data from subjects at risk for Alzheimer’s disease, we show that the sensitivity and power of statistical analysis of PIB-PET images can be improved by using the proposed method. The code will be made publicly available.

6. Acknowledgment

This research was supported by NIH grants AG040396, AG021155, AG037639, and NSF CAREER award 1252725. Partial support was provided by UW ADRC (AG033514), UW ICTR (1UL1RR025011), UW CPCP (AI117924) and NIH grants AG010129 and AG027161.

References

- [1] M. Aubry, U. Schlickewei, and D. Cremers. The wave kernel signature: A quantum mechanical approach to shape analysis. In *ICCV Workshops*, pages 1626–1633. IEEE, 2011. 5
- [2] M. Belkin and P. Niyogi. Laplacian Eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):pp. 1373–1396, 2003. 3
- [3] C. Brislawn. Traceable integral kernels on countably generated measure spaces. *Pacific Journal of Mathematics*, 150(2):229–240, 1991. 4
- [4] M. M. Bronstein and I. Kokkinos. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *CVPR*, pages 1704–1711. IEEE, 2010. 5
- [5] D. Chan, N. C. Fox, R. I. Scahill, et al. Patterns of temporal lobe atrophy in semantic dementia and Alzheimer’s disease. *Annals of Neurology*, 49(4):433–442, 2001. 8
- [6] R. R. Coifman and M. J. Hirn. Diffusion maps for changing data. *Applied and Computational Harmonic Analysis*, 36(1):79–107, 2014. 2, 4
- [7] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006. 2, 3, 4
- [8] M. T. Eismann, J. Meola, and R. C. Hardie. Hyperspectral change detection in the presence of diurnal and seasonal variations. *IEEE Geoscience and Remote Sensing*, 46(1):237–249, 2008. 2
- [9] Y. Fang, M. Sun, M. Kim, and K. Ramani. Heat-mapping: A robust approach toward perceptually consistent mesh segmentation. In *CVPR*, pages 2145–2152. IEEE, 2011. 5
- [10] K. J. Friston. Statistical parametric mapping. In *Neuroscience Databases*, pages 237–250. Springer, 2003. 1
- [11] D. Hammond, P. Vandergheynst, and R. Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011. 3, 8
- [12] H. Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *CVPR*, volume 2, pages 807–814. IEEE, 2005. 2
- [13] H. Hirschmuller. Stereo processing by semiglobal matching and mutual information. *IEEE PAMI*, 30(2):328–341, 2008. 2
- [14] J. E. Iglesias, E. Konukoglu, D. Zikic, et al. Is synthesizing MRI contrast useful for inter-modality analysis? In *MICCAI*, pages 631–638. Springer, 2013. 2
- [15] L. M. Ittner and J. Götz. Amyloid- β and tau – a toxic pas de deux in Alzheimer’s disease. *Nature Reviews Neuroscience*, 12(2):67–72, 2010. 7
- [16] J. Kim, V. Kolmogorov, and R. Zabih. Visual correspondence using energy minimization and mutual information. In *ICCV*, pages 1033–1040. IEEE, 2003. 2
- [17] W. H. Kim, M. K. Chung, and V. Singh. Multi-resolution shape analysis via Non-euclidean wavelets: Applications to mesh segmentation and surface alignment problems. In *CVPR*, pages 2139–2146. IEEE, 2013. 5
- [18] W. H. Kim, D. Pachauri, C. Hatt, et al. Wavelet based multi-scale shape features on arbitrary surfaces for cortical thickness discrimination. In *NIPS*, pages 1250–1258, 2012. 8
- [19] W. H. Kim, V. Singh, M. K. Chung, et al. Multi-resolutional shape features via non-Euclidean wavelets: Applications to statistical analysis of cortical thickness. *NeuroImage*, 93:107–123, 2014. 8
- [20] W. E. Klunk, H. Engler, A. Nordberg, et al. Imaging brain amyloid in Alzheimer’s disease with Pittsburgh Compound-B. *Annals of neurology*, 55(3):306–319, 2004. 2
- [21] F. LaFerla. Amyloid- β and tau in Alzheimers disease. *Nature Reviews Neuroscience*, 2008. 7
- [22] F. Maes, A. Collignon, D. Vandermeulen, et al. Multimodality image registration by maximization of mutual information. *IEEE TMI*, 16(2):187–198, 1997. 2
- [23] S. Mallat. *A wavelet tour of signal processing*. Academic press, 1999. 3
- [24] Y. Notay. Combination of Jacobi–Davidson and conjugate gradients for the partial symmetric eigenproblem. *Numerical Linear Algebra with Applications*, 9(1):21–44, 2002. 7
- [25] N. F. Osman and J. L. Prince. Regenerating MR tagged images using harmonic phase (harp) methods. *IEEE Biomedical Engineering*, 51(8):1428–1433, 2004. 2
- [26] J. P. Pluim, J. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE TMI*, 22(8):986–1004, 2003. 2
- [27] E. M. Reiman, K. Chen, X. Liu, et al. Fibrillar amyloid- β burden in cognitively normal people at 3 levels of genetic risk for Alzheimer’s disease. *Proceedings of the National Academy of Sciences*, 106(16):6820–6825, 2009. 8
- [28] S. Roy, A. Carass, N. Shiee, et al. MR contrast synthesis for lesion segmentation. In *Biomedical Imaging: From Nano to Macro*, pages 932–935. IEEE, 2010. 2
- [29] P. Scheltens, D. Leys, F. Barkhof, et al. Atrophy of medial temporal lobes on MRI in probable Alzheimer’s disease and normal ageing: diagnostic value and neuropsychological correlates. *Journal of Neurology, Neurosurgery & Psychiatry*, 55(10):967–972, 1992. 8
- [30] S. Haykin and B. V. Veen. *Signals and Systems, 2nd Edition*. Wiley, 2005. 3
- [31] M. Sun, Y. Fang, and K. Ramani. Center-shift: an approach towards automatic robust mesh segmentation (arms). In *CVPR*, pages 630–637. IEEE, 2012. 5
- [32] P. Viola and W. M. Wells III. Alignment by maximization of mutual information. *ICCV*, 24(2):137–154, 1997. 2
- [33] W. M. Wells III, P. Viola, H. Atsumi, et al. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, 1996. 2