

# On Statistical Analysis of Neuroimages with Imperfect Registration

Won Hwa Kim<sup>†‡</sup> Sathya N. Ravi<sup>\*</sup> Sterling C. Johnson<sup>\*‡</sup> Ozioma C. Okonkwo<sup>\*‡</sup> Vikas Singh<sup>§†‡</sup>

<sup>†</sup>Dept. of Computer Sciences, University of Wisconsin, Madison, WI

<sup>§</sup>Dept. of Biostatistics & Med. Informatics, University of Wisconsin, Madison, WI

<sup>\*</sup>Dept. of Industrial and Systems Engineering, University of Wisconsin, Madison, WI

<sup>‡</sup>Wisconsin Alzheimer's Disease Research Center, University of Wisconsin, Madison, WI

<sup>\*</sup>GRECC, William S. Middleton VA Hospital, Madison, WI

## Abstract

A variety of studies in neuroscience/neuroimaging seek to perform statistical inference on the acquired brain image scans for diagnosis as well as understanding the pathological manifestation of diseases. To do so, an important first step is to register (or co-register) all of the image data into a common coordinate system. This permits meaningful comparison of the intensities at each voxel across groups (e.g., diseased versus healthy) to evaluate the effects of the disease and/or use machine learning algorithms in a subsequent step. But errors in the underlying registration make this problematic, they either decrease the statistical power or make the follow-up inference tasks less effective/accurate. In this paper, we derive a novel algorithm which offers immunity to local errors in the underlying deformation field obtained from registration procedures. By deriving a deformation invariant representation of the image, the downstream analysis can be made more robust as if one had access to a (hypothetical) far superior registration procedure. Our algorithm is based on recent work on scattering transform. Using this as a starting point, we show how results from harmonic analysis (especially, non-Euclidean wavelets) yields strategies for designing deformation and additive noise invariant representations of large 3-D brain image volumes. We present a set of results on synthetic and real brain images where we achieve robust statistical analysis even in the presence of substantial deformation errors; here, standard analysis procedures significantly under-perform and fail to identify the true signal.

## 1. Introduction

A broad spectrum of research in neuroscience including studies focused on understanding the process of aging, the effects of trauma and the manifestation of neurological disorders rely heavily on statistical analysis of neuroimaging data [26]. Typical questions may include identifying which

brain regions are affected as a function of disease and/or finding associations or correlations of regional measurements in the image with future cognitive decline [10, 16, 8]. A key component which precedes much of the “analysis” pipeline is co-registration [1]. For instance, if the study cohort includes a set of 100 participants/subjects, the co-registration step will warp each of the 100 images into a common *template* coordinate system. It is easy to see that doing so enables one to retrieve the measurement at a specific voxel  $v$  across the entire cohort — in other words, for any arbitrary voxel in one image, we know its corresponding voxels in all other images. This process, often called as “spatial normalization”, is essential for performing either voxel-wise inference (i.e., independently, one voxel at a time) or utilizing such data for classification or regression tasks [29, 35, 36].

Given the critical role of registration in statistical image analysis of brain imaging data, the community has invested much effort in algorithms and (mature) software libraries that are reliable and easily deployable. Depending on the clinical question and the type of image modality under study, methods based on non-linear/free-form deformations optimizing various loss functions (e.g., mutual information) are broadly available [20, 21, 27, 24]. Note that while the general ideas driving image registration do not change very frequently, most popular toolboxes are regularly fine-tuned, to achieve better accuracy. These “incremental” refinements seem routine, but are very valuable in practice. To see why, let us use an example to answer a closely related question, namely, what happens when the registration pre-processing is slightly imperfect?

Consider a dataset which consists of two groups: healthy and diseased subjects. Once the initial co-registration is done, our goal may be to perform a voxel-wise parametric test to identify voxels that are statistically different across the groups suggesting a potential disease effect [2]. It is commonly observed by practitioners that using a newer registration method that is marginally better relative to its older

version improves the final analysis in measurable ways: for example, by revealing a *stronger* disease effect (i.e., *lower p-values*) and possibly more image regions where differences are statistically significant (survive a threshold of  $\alpha = 0.05$ ) across the healthy and diseased groups. This suggests that even small errors in the registration may have a detrimental effect on the downstream analysis. The statistical power improvements described above have important and real consequences; an actual experiment or study that may have needed 500 participants to evaluate a hypothesis may now be possible (using a better registration procedure) with a slightly smaller sample size.

Clearly, image registration algorithms will continue to improve. But separate from these developments, this paper considers a complementary (but potentially more interesting) question that is *not tied to* which method is currently the best. Assume that there is an ideal registration (or transformation),  $\mathcal{T}$  and the one estimated by a state of the art registration method is  $\mathcal{T} + \epsilon$  with error  $\epsilon > 0$ . Based on the foregoing discussion, actively pursuing better registration schemes is obviously important. But independent of these improvements, can we derive statistical analysis methods that are, by design, *immune* to nominal values of  $\epsilon$ ? In other words, as long as the registration procedure provides a reasonable estimate of  $\mathcal{T}$ , the follow-up analysis operates on alternate representations of the image that are *invariant* to local deformations (of up to magnitude  $\epsilon' \leq \epsilon$ ) [17, 19, 25, 18, 34, 33]. Such a tool, if available, will have two direct implications. Notice that if the upstream registration is already good, such a local-deformation invariant analysis framework may occasionally offer a small improvement or at worst, will match the results that we obtained anyway. But more importantly, *if* the results from state-of-art registration methods are imperfect, such a framework will be extremely useful. In some sense, the final analysis results will be consistent with a (hypothetical) registration procedure that did *not* make those errors. Note that while spatial smoothing may provide some resilience to such local registration errors, it will be desirable to obtain algorithms that are invariant to such errors.

The **contribution** of this paper is to take a small step towards registration invariant analysis of brain imaging data. To do so, we leverage recent results from harmonic analysis, namely, scattering coefficients, to derive image representations that are provably invariant to local deformations. This so-called scatter transform is obtained via a cascade of operations, involving wavelet expansions using an orthonormal basis derived from a function of the image. The downstream statistical questions can then be simply reformulated in terms of such representations, which are immune to nominal levels of errors in the given registration. We show simulation results as well as empirical evidence obtained from experiments on real brain images.

## 2. The Continuous Wavelet Transform

The core of our proposed ideas relies on the theory of wavelets and their various applications in computer vision and image processing. Our framework will essentially apply these transforms in a sequential/network manner to the given data and as we will see shortly, the resultant coefficients when appropriately averaged will offer the type of invariance properties we desire. Since the wavelet expansion is critical to the description of our ideas, we will give a brief overview next and then describe our algorithm.

The wavelet transform is conceptually similar to the Fourier transform, in that a given function is transformed using a specific set of basis functions. While the Fourier transform uses  $\sin()$  bases with infinite support, the wavelet transform uses a mother wavelet basis  $\psi$  which is localized in both time and frequency domain [22].

The definition of wavelet transform starts from the construction of a mother wavelet  $\psi_{t,a}$ , a function of two parameters, a scale parameter  $t$  and a translation parameter  $a$ ,

$$\psi_{t,a}(x) = \frac{1}{t} \psi\left(\frac{x-a}{t}\right) \quad (1)$$

Here, the parameters  $t$  and  $a$  control the dilation and localization of  $\psi_{t,a}$  respectively. Using  $\psi_{t,a}$  as the bases, the wavelet transform of a function  $f(x)$  is defined as the inner product of the  $\psi$  and  $f$ , represented as

$$W_f(t, a) = \langle f, \psi \rangle = \frac{1}{t} \int f(x) \psi^*\left(\frac{x-a}{t}\right) dx \quad (2)$$

where  $W_f(t, a)$  is the wavelet coefficient at scale  $t$  and location  $a$ , whereas  $\psi^*$  is the complex conjugate of  $\psi$ .

For our purposes, it will in fact be more convenient to think of this transform as an operator  $T^t$  acting on  $f$ . We can check that this operation can now be represented in terms of convolution by defining  $\bar{\psi}_t(x) = \frac{1}{t} \psi^*\left(\frac{-x}{t}\right)$ . The expression in (2) becomes

$$(T^t f)(a) = \int \bar{\psi}_t(a-x) f(x) dx = (\bar{\psi}_t \star f)(a) \quad (3)$$

When the mother wavelets  $\psi_t$  ( $t = 1, 2, \dots, J$ ) of multiple scales are transformed to the frequency domain, they form a group of band-pass filters. Since these band-pass filters do *not* cover the low-frequency band, we need to introduce an additional low-pass filter, a scaling function  $\phi$ . Since the scaling function  $\phi$  acts as a low-pass filter, the forward transform using  $\phi$  returns a smoothed (or averaged) representation of the original function  $f$  in terms of the wavelet coefficients. The transform using  $\phi$  is often referred as the wavelet transform at scale 0 using  $\psi_{t_0}$ .

## 3. Obtaining a Deformation Invariant Representation via Scattering Coefficients

Our primary strategy is to use the recently proposed scattering transform of an image to obtain a deforma-

tion invariant representation. Consider a set  $I = \{I_1, \dots, I_N, I_{N+1}, \dots, I_{2N}\}$  of images that we seek to perform classification experiments on, where the indices  $\{1, \dots, N\}$  and  $\{N + 1, \dots, 2N\}$  correspond to the first (and second) class respectively. If each of the images corresponds to handwritten digits with arbitrary orientations and translations, the scattering transform attempts to offer resilience to such variations. To do so, the scattering transform maps an image to a lower dimensional space where the distances between data classes are maximized (while within class distances are minimized). It does so *without* using the class labels and in this sense, is fully unsupervised. The underlying machinery uses a combination of rotational wavelets (i.e., Gabor filter) and modulus/averaging operations to obtain an embedding for the given image, to subsequently compute meaningful similarities between them despite the fact that the rotation/translation of each individual image is arbitrary. Due to some of these interesting properties, the scattering transform will serve as an excellent starting point for our algorithm which we will then modify to derive the type of behavior needed in our application.

### 3.1. Scattering Transform

The scattering transform is motivated by a need for a representation that is robust to group actions. It utilizes the wavelet transform and the modulus operator together to obtain such group invariants. We know that the mother wavelet function  $\psi$  is shaped as a localized oscillating function with finite duration, which is different from the Fourier basis with infinite support. It is precisely this localization property that offers at least partial robustness to deformation within the classical wavelet transform. Unfortunately, it is this localization property that makes the wavelet transform *variant* to local translations, in fact, the wavelet transform is covariant. Recall that our objective is to perform image analysis even when the overall registration has not adequately estimated the local deformations. In free-form non-linear registration, local deformations are generally represented as local translations, which in addition to the covariance property of the wavelet transform poses a significant technical issue. To overcome this, the scattering transform uses an additional non-linear operation to make the resultant representations invariant to such localization which we will describe next.

There are several non-linear operations one can use in the above setting. One of the simplest non-linear operations that are popular and easy to analyze are piecewise linear operators. Scattering transform uses the modulus operator,  $M$ , which is in fact a piecewise linear operator and moreover it satisfies the following three important properties, each of which is important in a special way. It is pointwise, non-expansive (i.e.,  $\|Mf - Mg\| \leq \|f - g\|$ ) and signal energy preserving [3].

Now, we will see intuitively why scattering transform is

rotation invariant. In practice, scattering transform uses a finite number of filters at different phase angles. These phase angles are generally chosen to be the  $n$ -th root of unity and hence form a finite group under multiplication as the group operation. In order for us to come up with an invariance to this group (rotation) action, we can use a popular trick in group representation theory (see chapter 2 in [28]), specifically, to take the average over all the filters. We will see how this is done more explicitly in the next section. Note that even though the averaging operation makes sense only when the group is finite, one can nonetheless extend this line of thought to compact Lie groups by replacing the sum in the averaging operation by integral, please see [23] for more details.

To summarize the above discussion, the scattering transform obtains its non-linear invariants by combining wavelets (for additive noise invariance), modulus (for local translation invariance) and averaging operation (for local rotation or some other group invariance) in a cascade fashion thus obtaining a local deformation invariance. Let  $U[s]f = |f \star \psi_s|$ , and any sequence  $p = (s_1, s_2, \dots, s_m)$  define a path of an ordered product of operators as

$$U[p]f = U[s_m] \cdots U[s_2]U[s_1]f \quad (4)$$

$$= |\cdots ||f \star \psi_{s_1}| \star \psi_{s_2}| \cdots \star \psi_{s_1}| \quad (5)$$

Using this concept, a scattering transform (represented as a convolution network) along path  $p$  at the location  $u$  is defined as

$$S[p]f = U[p]x \star \phi(u) \quad (6)$$

$$= |\cdots ||f \star \psi_{s_1}| \star \psi_{s_2}| \cdots \star \psi_{s_1}| \star \phi(u) \quad (7)$$

yielding scattering coefficients  $S[p]f$  for path  $p$ . These scattering coefficients are invariant to translation of  $f$  and although it has many similarities with the Fourier transform modulus,  $S[p]f$  is Lipschitz continuous to deformation in contrast to the Fourier transform modulus. So far, applications of the scattering transform have been fairly limited to analyzing signals in 1-D and (small sized) 2-D settings and to a few problems such as texture [3, 31, 32] and hand writing analysis [3].

Despite these useful properties of the scattering transform, the standard construction of the scattering transform is inapplicable to our application where the dimension of the function space and size of the data is much larger. Many medical imaging modalities provide high dimensional images, and the number of scattering coefficients becomes very large and makes further analysis very inefficient, often infeasible. For instance, utilizing rotation filters at small incremental resolutions at all possible locations is computationally intractable for most image datasets of interest, especially 3-D images.

*Key Idea.* A potential solution to the above issue must utilize the underlying structure of the given data and avoid

a computational cost that increases rapidly with the dimension. The most natural mathematical object that offers such properties is a graph. In most medical images, each voxel is highly correlated to its spatial neighbors as well as across subjects for a given voxel location. When expressed as a graph, the correlations are encoded as edges whereas the number of nodes only depends on the size of the image, not on its dimensionality. In this scenario, intuitively, the notion of rotation can be substituted by strategies based on anisotropic filtering governed by the edge weights. If some additional technical conditions hold (to be described shortly), one can still offer the types of properties obtained in the conventional scattering transform.

While the above idea is promising, there is still one significant gap. Recall that the core part of the scattering transform depends on the wavelet transform, which classically has been defined *only* in the Euclidean space. In order to define scattering transform on data that lives in the non-Euclidean space (i.e., a graph), we first need to write down the wavelet expansion on a graph. Luckily, recent results from harmonic analysis provide theoretically rigorous constructions to do so [9, 11, 6]. Based on these ideas, we describe how a wavelet can be constructed and used to define a wavelet transformation in a graph setting, where in general, the notion of ‘scale’ and ‘translation’ in the spatial domain, essential in wavelets, are difficult to specify directly.

### 3.2. Wavelet Transformation on Graphs

A graph  $G = \{V, E, \omega\}$  consists of a vertex set  $V$ , an edge set  $E$  and corresponding edge weights  $\omega$ , and let  $N$  be the number of vertices. The graph is given as an adjacency matrix  $A$ , where each element  $a_{ij}$  in  $A$  denotes the connection between vertex  $i$  and  $j$ . Here, vertices will correspond to distinct voxels in the image. The degree matrix  $D$  of  $G$ , is a diagonal matrix where the  $i$ th diagonal is the sum of all the edge weights associated with the vertex  $i$ . From these two graph representation matrices, a graph Laplacian is computed as  $L = D - A$ . The spectrum of the graph Laplacian  $L$  yields the eigenvalues  $\lambda_l$  and corresponding the eigenfunctions  $\chi_l, l = 0, 1, \dots, N - 1$ , which defines a dual domain analogous to the frequency domain for  $\mathbb{R}^n$ .

Invoking spectral graph theory [5] offers a very important advantage in this setting. Since the graph Laplacian  $L$  is a self-adjoint operator, one can use these  $\chi_l$  as the bases to define the graph Fourier transform as

$$\hat{f}(l) = \sum_{n=1}^N \chi_l^*(n) f(n) \text{ and } f(n) = \sum_{l=0}^{N-1} \hat{f}(l) \chi_l(n) \quad (8)$$

where  $\hat{f}$  is the transformed function in the frequency domain. Notice that the only difference between the formulation above and the textbook definition of the Fourier transform is the bases we use, which in this case are derived from a self-adjoint operator of the graph. Interestingly, the

above expression will be used next to perform the Wavelet transform of a graph.

As mentioned in Section 2, wavelets in the frequency domain serve as band-pass filters. Therefore, in order to construct a mother wavelet, the intuition is to design a band-pass filter function  $g()$  in the frequency domain instead of implementing one in the original domain. The mother wavelet  $\psi$  can be constructed by first defining  $g$  in the frequency domain and then localizing it by a delta function  $\delta$  in the original domain. Utilizing the graph Fourier transform and the idea above, a mother wavelet  $\psi_{s,n}$  at scale  $s$  and localized at a vertex  $n$  is implemented as

$$\psi_{t,n}(m) = T_g^t \delta(n) = \sum_{l=0}^{N-1} g(t\lambda_l) \chi_l^*(n) \chi_l(m) \quad (9)$$

where scale  $t$  is defined inside the kernel function  $g$  using the scaling property of Fourier transform [30]. Notice that the eigenvalues serve the analog of the frequency in this setting. Once the bases  $\psi$  are constructed, the wavelet transform operator at scale  $s$  on function  $f$  is defined as

$$T_g^t f(n) = W_f(t, n) = \langle f, \psi_{t,n} \rangle \quad (10)$$

$$= \sum_{l=0}^{N-1} g(t\lambda_l) \hat{f}(l) \chi_l(n) \quad (11)$$

Such a wavelet transform on graphs, known as the recently proposed Spectral Graph Wavelet Transform (SGWT) [9], is not novel to this paper but has been limited to only few problems in vision and machine learning [12, 13, 14].

### 3.3. Scattering Transform on Graphs

With the above components in hand, deriving a Scattering transform on a graph can be performed mechanically. First, we define a single scattering operation by combining the wavelet and modulus operations together as,

$$S(t, n) = |f \star \bar{\psi}_{t,n}| \quad (12)$$

which yields the local deformation invariant coefficients  $S(t, n)$  in scale  $t$  at vertex  $n$ . Note that in order for a transform to achieve robustness to deformation, *it must be non-expansive*, so we should expect the graph scattering transform presented above to offer these properties. We provide this analysis next.

First, notice that  $S(\cdot, \cdot)$  is local deformation invariant since  $\bar{\psi}$  is one. Now, we show that it is also invariant to additive noise (possibly as a consequence of incorrect local warps and subsequent interpolations) which is an important property that does not hold for the classical Fourier transform. This can be accomplished by showing that it is a non-expansive operator [23],

**Proposition 1.** *The scattering operation  $S$  on a graph using SGWT operator  $T_g^s$  at scale  $s$  with a kernel  $g()$  that is  $\int g(x) dx = 1$  is non-expansive.*

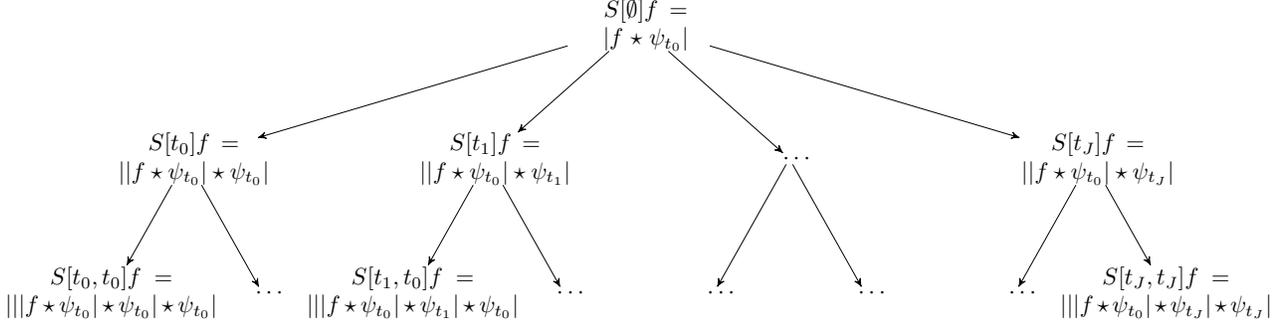


Figure 1: Illustration of the convolution network using scattering operation. A function  $f$  propagated through the network using SGWT and modulus operation. At first,  $f$  is averaged using scaling function  $\psi_{t_0}$  and modulus transformed to obtain the scattering coefficients in the first layer. The scattering coefficients are recursively propagated through the branches by convolution using SGWT in multiple scales, then applied with modulus operation to generate scattering coefficients in the next layer.

*Proof.* Given two functions  $f$  and  $h$  defined on a graph,

$$\begin{aligned}
\|Sf - Sh\| &= \left\| |T_g^s f| - |T_g^s h| \right\| \\
&= \left\| \left| \sum_{l=0}^{N-1} g(t\lambda_l) \hat{f}(l) \chi_l \right| - \left| \sum_{l=0}^{N-1} g(t\lambda_l) \hat{h}(l) \chi_l \right| \right\| \\
&\leq \left\| \sum_{l=0}^{N-1} g(t\lambda_l) \hat{f}(l) \chi_l - \sum_{l=0}^{N-1} g(t\lambda_l) \hat{h}(l) \chi_l \right\| \\
&= \|f - h\| \quad \square
\end{aligned}$$

We can now construct the scattering operation in a cascade fashion in multiple layers, and obtain a convolutional network that provides a deformation invariant representation of a given function  $f$  as

$$S[\emptyset]f = |f * \bar{\psi}_{t_0}| \quad (13)$$

$$S[t_{l_1}]f = ||f * \bar{\psi}_{t_0}| * \bar{\psi}_{t_{l_1}}| \quad (14)$$

$$S[t_{l_1}, t_{l_2}]f = |||f * \bar{\psi}_{t_0}| * \bar{\psi}_{t_{l_1}}| * \bar{\psi}_{t_{l_2}}| \quad (15)$$

⋮

where  $S[l]f$  gives the scattering coefficients obtained by the wavelet bases at the  $l$ th layer. In layer  $l_k$ , the scattering coefficients from the previous layer  $l_{k-1}$  are processed by the forward wavelet transform at multiple scales and then the modulus operator is applied to generate the new scattering coefficients to be passed to the next layer  $l_{k+1}$ .

Our construction of scattering transform is distinct from the transform proposed in [3] in that we use scales, not the rotation, for branching. Notice that in a graph setting, we do not have a notion of direction, but the anisotropic filtering is still affected by the graph edge weights. By branching out using scales instead of rotation, we obtain a multi-resolution view of the scattering coefficients from each layer and local stability to scales. Additionally, the use of spectral graph theory to derive the scattering transform is new and likely interesting for various vision applications. The work in [4] explores a related idea but it uses a different construction of scattering transform and experiments are demonstrated on smaller digits datasets.

## 4. Experimental Result

In this section, we demonstrate two sets of experimental results focused on applying a standard statistical group difference analysis pipeline: on the original representation of the data and then, the deformation invariant representations derived by our method. We describe experiments both with synthetic images and real brain image datasets. First, we assess group differences between two groups of simulated images where the goal is to detect the *true* voxel-wise group-level differences between the first and the second group, even when the images undergo arbitrary local rotations and deformations as well as additive noise in each image sample in the cohort. When the registration is good, the analysis will pick out voxels where the distributions of measurements (intensities) are statistically different across the two groups. Next, we evaluate group differences based on a well-known risk factor for Alzheimer’s disease (AD) on a cohort of real fluorodeoxyglucose Positron emission tomography (FDG-PET) images. Here, when the images are perfectly registered, the statistical group analysis results are evident — those regions known to be affected by the disease/risk factor are identified. However, if there are errors in the underlying deformation field obtained from the registration (quite common in practice), the analysis loses its statistical power because we are not comparing corresponding voxels across images. With various levels of errors, the results progressively deteriorate. Ideally, if the proposed algorithm works perfectly, we will obtain group differences as if we had an ‘error-free’ registration procedure.

### 4.1. Toy Example for Group Analysis of Images with Rotation and Translation

In this section, we demonstrate results of statistical analysis for identifying group level differences using our framework on a population of synthetic images, representative of two distinct groups.

**Design.** For the first group, we consider a default image of an apple (a proxy for the first group) and for the second group, we artificially introduce two separate holes,

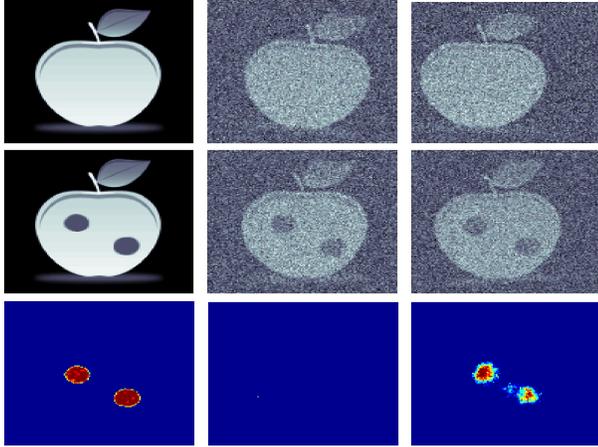


Figure 2: Toy example of our framework on a dataset of images from two groups. First row: the mean image (left) and synthetic images (middle and right) from the first group, Second row: the mean image (left) and synthetic images (middle and right) from the second group, Third row:  $p$ -value maps from group analysis using initial images (left), using synthetic images (middle) and using our framework (right). The bottom row shows that standard statistical analysis detects the group differences correctly on registered images, but it fails in presence of rotation and translation in the images. However, our algorithm accurately detects the group differences (the two holes) despite the deformations.

which are representatives of the group-level differences, see top two rows of the first column in Fig. 2. Based on these ‘models’ for the two groups, which serve as the mean  $\mu$ , we draw samples from each group to populate the cohort: 20 images for each group (total of 40). The image intensities are normalized to  $[0, 1]$ . In other words, for each individual image, the intensity at a specific pixel is drawn from a distribution centered on the mean intensity for *that pixel* in the model image (first or second group). At this stage, the images are in correspondence and performing a pixel-wise parametric test is meaningful and will reveal precisely the holes as the group-wise difference. Now, we apply a transform  $A$ , which consists of random rotation and translation drawn from Gaussian distribution (with a mean of 5 percent of the image space for translation,  $10^\circ$  for rotation and a variance 1 for both), to each image to simulate deformations in the images. The pixels across the dataset are *no longer in correspondence*, so a pixel-wise test cannot be performed. We then add Gaussian noise to each pixel in each image. The modified images for the first and second groups are,

$$Y_{\text{group } 1}(i, j) = A\mu_{\text{group } 1}(i, j) + N(0, 0.3) \quad (16)$$

where  $A$  is an arbitrary transformation sampled from a distribution described above. Examples of these synthetic images are shown in the top two rows of the second and the third column of Fig. 2.

From the two representative images (top and middle row images in the first column of Fig. 2), we can easily tell that the true difference between the models of the groups are the two holes. When standard statistical group analysis is applied pixel by pixel on the two groups of images (with noise

and transformation), by performing a  $t$ -test at each pixel and using Bonferroni correction at  $\alpha = 0.05$  level for multiple comparisons correction, we detect no pixels as showing significant group differences. This is expected because the pixels are no longer in correspondence. On the other hand, using our framework, i.e., first constructing grid graph using the image intensities, applying scattering transform on the graphs to obtain scattering coefficients and performing a hypothesis test at each pixel, and finally, applying a Bonferroni correction at  $\alpha = 0.05$  level, we are able to detect most of the true group differences. This result demonstrates that our framework does offer a reasonable level of invariance to local rotation and translations. If we imagine the two holes in the second group is a consequence of some pathology and the samples drawn from that distribution reflect individual level variations, these simulations suggest that even if the registration is not perfect, the downstream statistical analysis can be made robust to such errors.

## 4.2. Statistical Analysis on FDG-PET Scans

In this study, we utilize a general statistical group analysis pipeline on real 3-D FDG-PET scans to identify those brain regions that are related to Alzheimer’s disease (AD) risk factors such as the Apolipoprotein E (APOE) genotype of the subjects and evaluate our framework in the presence of varying levels of registration error. Note that an increase in FDG is a known manifestation of AD pathology, which is caused by an uptake of glucose in certain brain tissue. APOE  $\epsilon 4$  is genetically related to the development of AD, where the risk for AD is largely increased with increase in the number of APOE  $\epsilon 4$  alleles [7].

**Dataset.** We use a dataset of 130 healthy control participants where some subjects have potential AD risk factors (i.e., APOE  $\epsilon 4$  genotype, family history and etc). The cohort is comprised of 38 males and 92 females, with a mean age of 64.18. The 3-D FDG-PET scans are spatially registered to the Montreal Neurological Institute (MNI) space, and image intensity values are normalized using intensities from the cerebellum as the reference region. During the registration process, a deformation field — a warp field of the original image to the template image, is acquired. We introduce errors in this deformation field by 1) adding uncorrelated errors, up to 5%–10% of the deformation magnitude at each voxel, 2) adding realizations of a Gaussian

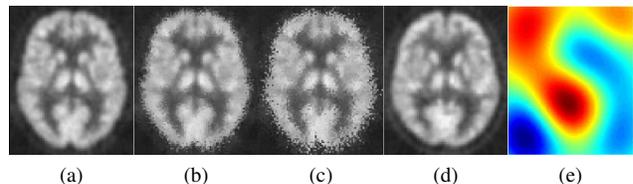


Figure 3: Registered FDG-PET scans of a subject. a) Using the original deformation field, b) Using deformation field with 5% noise level, c) Using deformation field with 10% noise level, d) Using deformation field with spatially correlated noise, e) A slice of GRF used for generating d).

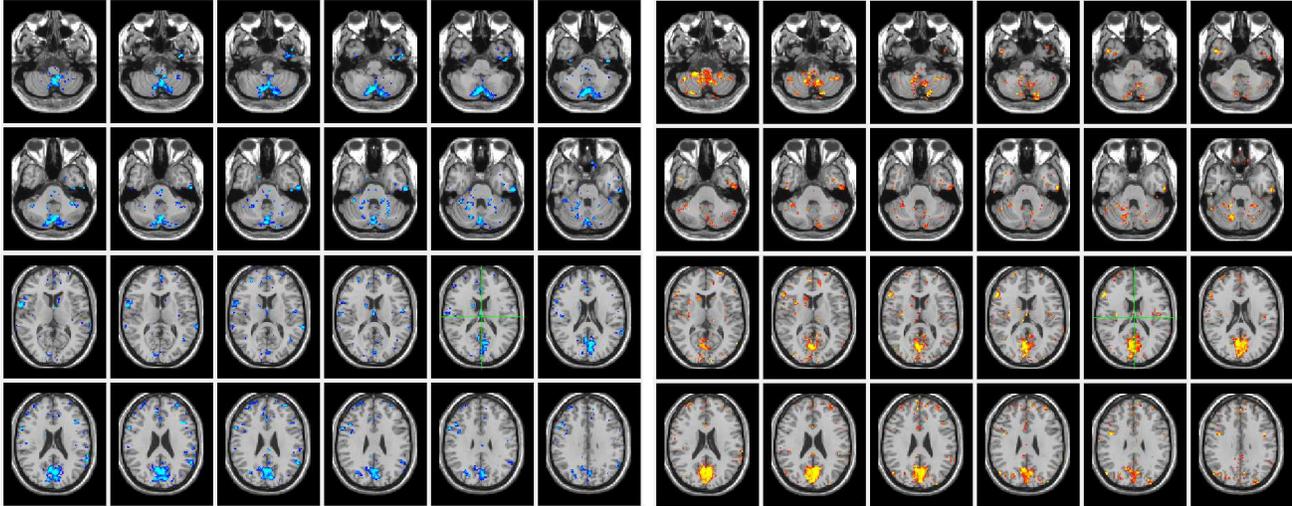


Figure 4: Montage of axial views of the  $p$ -value (in  $-\log_{10}$  scale) map on a template T1-weighted brain image. Group analysis was performed between APOE  $\epsilon 4$  genotype positive and negative groups. On the left column, the blue-light blue intensities indicate  $p$ -values using properly registered scans, and on the right column, the red-yellow intensities indicate  $p$ -value map using imperfect registration (5% noise level), both in the range of  $[0, 0.05]$ . We can see that the identified regions from both analysis are consistent though the second row shows a few regions which were not picked up by our algorithm.

Random Field (GRF) with variance 0.5 to simulate spatially correlated error ( $\sim 5$  voxels). Examples of these imperfectly registered images are shown in Fig. 3.

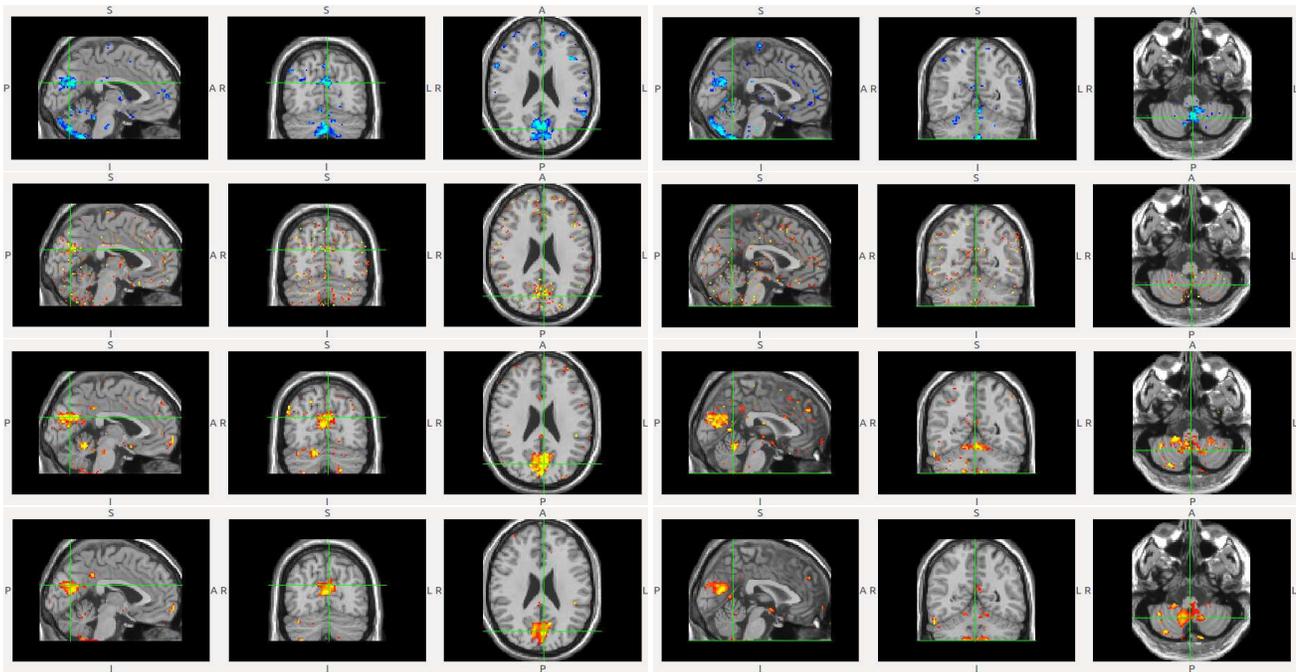
**Experimental Setup.** In this experiment, we used three different datasets with different noise levels in the deformation field for registration (Gaussian noise with mean of 0%, 5% and 10% of the warp magnitude and variance 1). As a baseline, we used properly registered FDG-PET images and divided them into two groups using APOE  $\epsilon 4$  status, a well-known AD risk factor, to find which brain regions are most strongly affected by the genotype. This baseline result will serve as the ground truth showing risk factor specific regions. We applied  $t$ -test at each voxel and thresholded the resultant  $p$ -values at 0.05 level. These  $p$ -values were projected onto a T1-weighted template image to identify the risk factor specific regions.

On the imperfectly registered datasets, i.e., 1) 5% and 10% noise levels, 2) GRF structured noise in the deformation field, we applied our framework to detect the risk factor specific brain regions. Given an FDG-PET image  $I$  for each subject, we first constructed a grid graph (i.e., defined each voxel as a vertex and defined an edge between a voxel and its six neighboring voxels in the 3-D volume space with edge weights of  $\exp(-\|I(x) - I(y)\|^2/\sigma^2)$ ). Then, the proposed scattering transform on graphs was applied on each grid graph to generate scattering coefficients. For the scattering transform, we used a wavelet transform with five scales for each layer, and we went up to the third layer of the cascade to derive scattering coefficients for the original image. In much of brain image analysis, lower frequency components with less noise are preferred, so we took the first three descendant branches from each branch in each

layer. Then, these coefficients were defined in a vector form at each voxel (total of 13 features), and Hotelling's  $T^2$  test (a generalized version of student  $t$ -test) was applied at each voxel. Again, these resultant  $p$ -values were thresholded at 0.05 level and projected onto a T1-weighted template image to obtain an apples-to-apples comparison with the baseline.

**Analysis.** The main result of our analysis is demonstrated in Fig. 4. The baseline analysis using APOE  $\epsilon 4$  as a predictor on properly registered FDG-PET images revealed cuneus regions, known to be closely tied to AD pathology [15], and some lower cerebellum regions as shown in blue-light blue regions in Fig. 4. Using our algorithm on the imperfectly registered scans (5% noise level), we also detected the exact same regions with almost the same statistical power as shown in red-yellow regions in Fig. 4. Note that when the general statistical procedure (baseline analysis), was applied on the imperfectly registered images, the statistical power deteriorates as the noise level increases, and it fails to detect the group differences.

A closer view of the regions identified in an orthogonal view are shown in Fig. 5. In the first row of Fig. 5, they clearly show the cuneus (and some partial precuneus) and cerebellum regions respectively from the baseline analysis (in blue-light blue) with  $p$ -values in the range of  $[0, 0.05]$  in  $-\log_{10}$  scale. However, as shown in the second row of Fig. 5 in red-yellow, the statistical power to detect risk factor related regions diminishes due to noise in the registration process. At the 10% noise level, it returned salt-and-pepper type result, completely failing to obtain any meaningful result. In contrast, using our algorithm on the both datasets with 10% level voxel-wise noise and GRF, we were able to successfully detect *exactly* the same regions (i.e., cuneus



**Figure 5:** Comparison of statistical group analysis results on cuneus (left column) and cerebellum (right column) using FDG-PET scans with respect to APOE genotype. Resultant  $p$ -value maps on a T1-weighted image template are demonstrated. First row: Using properly registered scans (baseline), Second row: Using imperfectly registered scans (10% noise level), Third row: Our results on imperfectly registered scans (10% noise level), Fourth row: Our results on imperfectly registered scans (GRF noise). Compared to the baseline result in the first row, the results in the second row shows decreased statistical power with increased error in the registration, and is unable to detect the cuneus and lower cerebellum. However, even when there are errors during the registration (voxel-wise noise or GRF), our algorithm correctly identifies the cuneus as shown in the third and fourth rows, consistent with the analysis performed on properly registered images as in the first row.

and lower cerebellum regions) that were found in the baseline analysis, as demonstrated in red-yellow regions in the third and fourth row of Fig. 5. Note that the scattering coefficients do not exhibit any "ensemble" behavior and therefore cannot exploit voxel-wise independent noise in the registration. The results above indicate that even when a good registration is unavailable, our framework may offer a robust solution by performing the same analysis using a deformation invariant representation of the original images. All reported results have been controlled for multiple comparisons correction.

## 5. Conclusion

It is well known that the statistical analysis of brain imaging data (including classification and regression experiments) can only proceed once the images have been warped into a common template space. But when individual subjects have significant atrophy or a pathology (resulting from tumor or vascular factors), the registration is imperfect. Separately, in some populations creating a common template itself may be difficult which leads to sub-optimal registration. But independent of where the registration errors come from, their effect on the downstream analysis can be significant and has serious implications on the success of the study. We provide an algorithm that derives local deformation invariant representations of the image. In practice,

this means that inference using the image data can proceed as if a (much superior) registration method were available.

Our method is based on the recently proposed Scattering transform which we adapt in interesting ways using ideas related to spectral graph wavelets in the harmonic analysis literature. The performance of the proposed approach does not depend on whether the noise in the warp field is i.i.d. or correlated. The scattering operator is provably invariant to the action of any compact Lie group on the measured signal [23]. We show that deriving wavelet expansions using the graph representation of the data (together with other modifications) makes the scatter transform a viable tool for analyzing large 3-D image datasets. The proposed ideas have direct applications in neuroimage analysis but are likely to be more broadly applicable in other computer vision problems where invariance to group actions (e.g., rotation, translation, diffeomorphism) is desired. Finally, supplementary explaining details of our framework can be found at <http://pages.cs.wisc.edu/~wonghwa>.

## 6. Acknowledgment

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

## References

- [1] J. Ashburner and K. Friston. Multimodal image coregistration and partitioning a unified framework. *Neuroimage*, 6(3):209–217, 1997. [1](#)
- [2] J. Ashburner and K. J. Friston. Voxel-based morphometry: the methods. *Neuroimage*, 11(6):805–821, 2000. [1](#)
- [3] J. Bruna and S. Mallat. Invariant scattering convolution networks. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(8):1872–1886, 2013. [3](#), [5](#)
- [4] X. Chen, X. Cheng, and S. Mallat. Unsupervised deep haar scattering on graphs. In *Advances in Neural Information Processing Systems*, pages 1709–1717, 2014. [5](#)
- [5] F. R. Chung. *Spectral graph theory*, volume 92. AMS Bookstore, 1997. [4](#)
- [6] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006. [4](#)
- [7] E. Corder, A. Saunders, W. Strittmatter, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer’s disease in late onset families. *Science*, 261(5123):921–923, 1993. [6](#)
- [8] C. D. Good, I. S. Johnsrude, J. Ashburner, et al. A voxel-based morphometric study of ageing in 465 normal adult human brains. In *Biomedical Imaging*, pages 16–pp. IEEE, 2002. [1](#)
- [9] D. Hammond, P. Vandergheynst, and R. Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129 – 150, 2011. [4](#)
- [10] K. Herholz, E. Salmon, D. Perani, et al. Discrimination between Alzheimer dementia and controls by automated analysis of multicenter fdg pet. *Neuroimage*, 17(1):302–316, 2002. [1](#)
- [11] T. Hou and H. Qin. Admissible diffusion wavelets and their applications in space-frequency processing. *Visualization and Computer Graphics, IEEE Trans. on*, 19(1):3–15, 2013. [4](#)
- [12] W. H. Kim, B. B. Bendlin, M. K. Chung, et al. Statistical inference models for image datasets with systematic variations. In *CVPR*, pages 4795–4803. IEEE, 2015. [4](#)
- [13] 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. [4](#)
- [14] 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. [4](#)
- [15] 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. [7](#)
- [16] 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. [1](#)
- [17] E. Kowalski. *An introduction to the Representation Theory of Groups*, volume 155. American Mathematical Society, 2014. [2](#)
- [18] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back. Face recognition: A convolutional neural-network approach. *Neural Networks, IEEE Transactions on*, 8(1):98–113, 1997. [2](#)
- [19] H. Ling and D. W. Jacobs. Deformation invariant image matching. In *ICCV*, volume 2, pages 1466–1473. IEEE, 2005. [2](#)
- [20] F. Maes, A. Collignon, D. Vandermeulen, et al. Multimodality image registration by maximization of mutual information. *IEEE TMI*, 16(2):187–198, 1997. [1](#)
- [21] J. A. Maintz and M. A. Viergever. A survey of medical image registration. *Medical image analysis*, 2(1):1–36, 1998. [1](#)
- [22] S. Mallat. *A wavelet tour of signal processing*. Academic press, 1999. [2](#)
- [23] S. Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012. [3](#), [4](#), [8](#)
- [24] D. Mattes, D. R. Haynor, H. Vesselle, et al. PET-CT image registration in the chest using free-form deformations. *IEEE TMI*, 22(1):120–128, 2003. [1](#)
- [25] J.-M. Morel and G. Yu. Asift: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2):438–469, 2009. [2](#)
- [26] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols. *Statistical parametric mapping: the analysis of functional brain images: the analysis of functional brain images*. Academic press, 2011. [1](#)
- [27] A. Roche, G. Malandain, X. Pennec, et al. The correlation ratio as a new similarity measure for multimodal image registration. In *MICCAI*, pages 1115–1124. Springer, 1998. [1](#)
- [28] J. P. Serre. *Linear representations of finite groups*. 1977. [3](#)
- [29] D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, et al. Magnetic resonance image tissue classification using a partial volume model. *NeuroImage*, 13(5):856–876, 2001. [1](#)
- [30] S. Haykin and B. V. Veen. *Signals and Systems, 2nd Edition*. Wiley, 2005. [4](#)
- [31] L. Sifre and S. Mallat. Combined scattering for rotation invariant texture analysis. In *European Symposium on Artificial Neural Networks*, 2012. [3](#)
- [32] L. Sifre and S. Mallat. Rotation, scaling and deformation invariant scattering for texture discrimination. In *CVPR*, pages 1233–1240. IEEE, 2013. [3](#)
- [33] I. J. Simpson, M. Woolrich, A. R. Groves, and J. A. Schnabel. Longitudinal brain mri analysis with uncertain registration. In *MICCAI*, pages 647–654. Springer, 2011. [2](#)
- [34] I. J. Simpson, M. W. Woolrich, J. L. Andersson, A. R. Groves, J. Schnabel, et al. Ensemble learning incorporating uncertain registration. *TMI*, 32(4):748–756, 2013. [2](#)
- [35] S. M. Smith, M. Jenkinson, M. W. Woolrich, et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, 23:S208–S219, 2004. [1](#)
- [36] D. Zhang, Y. Wang, L. Zhou, et al. Multimodal classification of Alzheimer’s disease and mild cognitive impairment. *Neuroimage*, 55(3):856–867, 2011. [1](#)