Deformation-Invariant Sparse Coding for Modeling Spatial Variability of Functional Patterns in the Brain

George H. Chen, Evelina G. Fedorenko, Nancy G. Kanwisher, and Polina Golland

Massachusetts Institute of Technology, Cambridge MA 02139, USA

Abstract. For a given cognitive task such as language processing, the location of corresponding functional regions in the brain may vary across subjects relative to anatomy. We present a probabilistic generative model that accounts for such variability as observed in fMRI data. We relate our approach to sparse coding that estimates a basis consisting of functional regions in the brain. Individual fMRI data is represented as a weighted sum of these functional regions that undergo deformations. We demonstrate the proposed method on a language fMRI study. Our method identified activation regions that agree with known literature on language processing and established correspondences among activation regions across subjects, producing more robust group-level effects than anatomical alignment alone.

1 Introduction

Spatial variability of activation patterns in the brain poses significant challenges to finding functional correspondences across subjects. This variability results in misalignment of individual subjects' activations in an anatomically-normalized space. Consequently, the standard approach of averaging activations in such a space for group analysis sometimes fails to identify functional regions that are spatially variable across individuals, e.g., regions for higher-order tasks such as language processing.

Recent work addresses this variability in different ways [5,6,9]. Thirion *et al.* [6] identify contiguous regions, or parcels, of functional activation at the subject level and then find parcel correspondences across subjects. While this approach yields reproducible activation regions and provides spatial correspondences across subjects, its bottom-up, rule-based nature does not incorporate a notion of a group template while finding the correspondences. Instead, it builds a group template as a post-processing step. As such, the model lacks a clear group-level interpretation of the estimated parcels. In contrast, Xu *et al.* [9] use a spatial point process in a hierarchical Bayesian model to describe functional activation regions. Their formulation accounts for variable shape of activation regions and has an intuitive interpretation of group-level activations. However, since the model represents shapes using Gaussian mixture models, functional regions of complex shape could require a large number of Gaussian components.

G. Langs et al. (Eds.): MLINI 2011, LNAI 7263, pp. 68–75, 2012.

Lastly, Sabuncu *et al.* [5] sidestep finding functional region correspondences altogether by estimating voxel-wise correspondences through groupwise registration of functional activation maps from different subjects. This approach does not explicitly model functional regions.

We propose a novel way to characterize spatial variability of functional regions that combines ideas from [5,6,9]. We model each subject's activation map as a weighted sum of group-level functional activation parcels that undergo a subject-specific deformation. Our contributions are twofold. First, similar to Xu *et al.* [9], we define a hierarchical generative model, but instead of using a Gaussian mixture model to represent shapes, we represent each parcel as an image, which allows for complex shapes. By explicitly modeling parcels, our model yields parcel correspondences across subjects, similar to [6]. Second, we assume that the template regions can deform to account for spatial variability of activation regions across subjects. This involves using groupwise registration similar to [5] that is guided by estimated group-level functional activation regions. We perform inference within the proposed model using a variant of the expectation-maximization (EM) algorithm [1] and illustrate our method on the language system, which is known to have significant functional variability [3].

2 Model

We let $\boldsymbol{I} = \{I_1, I_2, \ldots, I_N\}$ be the N observed images, $I_n \in \mathbb{R}^{|\Omega|}$ be the activation map for subject n ($1 \leq n \leq N$), and Ω be the set of voxels. We assume that a dictionary of K images $\boldsymbol{D} = \{D_1, D_2, \ldots, D_K\}$ generates the observed images $\boldsymbol{I}$. Importantly, each dictionary element corresponds to a group-level parcel. We treat the dictionary size K as a fixed constant; various model selection methods can be used to select K.

We assume that each observed image I_n is generated i.i.d. as follows. First, we draw weight vector $w_n \in \mathbb{R}^K$ where each scalar entry w_{nk} is independently sampled from distribution $p_w(\cdot; \theta_k)$. Then, we construct pre-image $J_n = \sum_{k=1}^{K} w_{nk} D_k$. The observed image $I_n = J_n \circ \Phi_n^{-1} + \varepsilon_n$ is the result of applying invertible deformation Φ_n to pre-image J_n and adding white Gaussian noise ε_n with variance σ^2. This process defines the joint probability distribution over the weight vector and observed image for a specific subject:

$$p(I_n, w_n | \Phi_n, \boldsymbol{D}; \boldsymbol{\theta}, \sigma^2) = \prod_{k=1}^{K} p_w(w_{nk}; \theta_k) \prod_{x \in \Omega} \mathcal{N}\left(I_n(x); \sum_{k=1}^{K} w_{nk} D_k(\Phi_n^{-1}(x)), \sigma^2 \right), \tag{1}$$

where $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_K\}$. We aim to infer dictionary $\boldsymbol{D}$ and each deformation Φ_n so that for future experiments performed on the same subjects, we can treat the dictionary and subject-specific deformations as fixed and just estimate the contribution of each dictionary element. By introducing group-level parcels as dictionary elements, the model implicitly contains parcel correspondences since it suffices to look at where a particular dictionary element appears in each subject. Furthermore, each subject need not have all dictionary elements present.

Model Parameters. We treat each deformation Φ_n as a random parameter with prior distribution $p_\Phi(\cdot)$, which can also be viewed as regularizing each deformation to prevent overfitting. Choice of the deformation prior allows us to leverage existing image registration algorithms. To prevent spatial drift of the dictionary elements, we constrain the average of the subject-specific deformations, defined as $\Phi_1 \circ \Phi_2 \circ \cdots \circ \Phi_N$, to be identity:

$$p(\boldsymbol{\Phi}) = \prod_{n=1}^{N} p_\Phi(\Phi_n) \cdot \mathbf{1}\{\Phi_1 \circ \cdots \circ \Phi_N = \mathrm{Id}\}\,, \tag{2}$$

where $\boldsymbol{\Phi} = \{\Phi_1, \ldots, \Phi_N\}$ and $\mathbf{1}\{\cdot\}$ is the indicator function that equals 1 when its argument is true and equals 0 otherwise.

We also treat each dictionary element D_k as a random parameter. To resolve an inherent degeneracy where scaling the intensity of dictionary element D_k by constant c and scaling w_{nk} by $1/c$ for all n results in the same observed images $\boldsymbol{I}$, we choose to constrain each dictionary element D_k to have bounded ℓ_2 norm: $\|D_k\|_2 \leq 1$. To encourage sparsity and smoothness, we introduce ℓ_1 and MRF penalties. To constrain each dictionary element to be a parcel, we require each D_k to be a contiguous, unimodal image. Formally,

$$p(D_k; \lambda, \gamma) \propto \exp\left\{-\lambda \sum_{x\in\Omega} |D_k(x)| - \frac{\gamma}{2} \sum_{x\in\Omega} \sum_{y\in\mathcal{N}(x)} (D_k(y) - D_k(x))^2\right\} \cdot \mathbf{1}\{D_k \text{ is a contiguous, unimodal image}\}\,, \tag{3}$$

where hyperparameters λ and γ are positive constants, and $\mathcal{N}(x)$ denotes the neighborhood of voxel x.

Other model parameters are treated as non-random: $\boldsymbol{\theta}$ parameterizes distribution $p_w(\cdot; \theta_k)$ for each k, and σ^2 is the variance of the Gaussian noise. We use MAP estimation for $\boldsymbol{D}$ and $\boldsymbol{\Phi}$ and ML estimation for $\boldsymbol{\theta}$ and σ^2 whereas hyperparameters λ and γ are currently hand-tuned.

For our experiments, we place independent exponential priors on each component of weight vector w_n and use diffeomorphic Demons registration [8] to estimate deformations $\boldsymbol{\Phi}$. In particular, we choose $p_w(w_{nk}; \theta_k) = \theta_k e^{-\theta_k w_{nk}}$, where $w_{nk} \geq 0$ and $\theta_k > 0$. Moreover, we define $p_\Phi(\Phi_n) \propto \exp\{-\mathrm{Reg}(\Phi_n)\}$, where $\mathrm{Reg}(\cdot)$ is the Demons registration regularization function [8]. Combining these distributions over weights $\boldsymbol{w} = \{w_1, \ldots, w_N\}$ and deformations $\boldsymbol{\Phi}$ with equations (1), (2), and (3), we obtain the full joint distribution:

$$\begin{aligned} &p(\boldsymbol{I}, \boldsymbol{w}, \boldsymbol{\Phi}, \boldsymbol{D}; \boldsymbol{\theta}, \sigma^2) \\ &\propto \prod_{n=1}^{N} \left\{ \exp\{-\mathrm{Reg}(\Phi_n)\} \prod_{k=1}^{K} e^{-\theta_k w_{nk}} \prod_{x\in\Omega} \mathcal{N}\left(I_n(x); \sum_{k=1}^{K} w_{nk} D_k(\Phi_n^{-1}(x)), \sigma^2 \right) \right\} \\ &\quad \cdot \prod_{k=1}^{K} \exp\left\{ -\lambda \sum_{x\in\Omega} |D_k(x)| - \frac{\gamma}{2} \sum_{x\in\Omega} \sum_{y\in\mathcal{N}(x)} (D_k(y) - D_k(x))^2 \right\}, \end{aligned} \tag{4}$$

where weights w_{nk}'s are non-negative, average deformation $\Phi_1 \circ \cdots \circ \Phi_N$ is identity, and each D_k is contiguous and unimodal with $\|D_k\|_2 \leq 1$.

Relation to Sparse Coding. With a heavy-tailed prior p_w concentrated at 0 and no deformations (i.e., deformations are identity), our model becomes equivalent to sparse coding [4]. We extend sparse coding by allowing dictionary elements to undergo subject-specific deformations. In contrast to previous dictionary learning approaches that assume perfect spatial correspondences (e.g., Varoquaux *et al.* [7]), we estimate jointly a set of deformations $\boldsymbol{\Phi}$ and the distribution for weight vectors $\boldsymbol{w}$ in addition to learning the dictionary elements. Effectively, we recover dictionary elements invariant to "small" deformations, where the "size" of a deformation is governed by the deformation prior.

Parameter Estimation. We use a variant[1] of the EM algorithm [1] to estimate model parameters $\boldsymbol{\Phi}, \boldsymbol{D}, \boldsymbol{\theta}, \sigma^2$. Derivations have been omitted due to space constraints. To make computation tractable, a key ingredient of the E-step is to approximate posterior distribution $p(\boldsymbol{w}|\boldsymbol{I}, \boldsymbol{\Phi}, \boldsymbol{D}; \boldsymbol{\theta}, \sigma^2)$ with a fully-factored distribution

$$q(\boldsymbol{w}; \boldsymbol{\mu}, \boldsymbol{\nu}) = \prod_{n=1}^{N} \prod_{k=1}^{K} \mathcal{N}^+(w_{nk}; \mu_{nk}, \nu_{nk}), \tag{5}$$

where $\mathcal{N}^+(\cdot; \mu_{nk}, \nu_{nk})$ is the probability density of the positive normal distribution (the normal distribution with mean μ_{nk} and variance ν_{nk} conditioned to have non-negative support). We let $(\hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{D}}, \hat{\boldsymbol{\theta}}, \hat{\sigma}^2, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\nu}})$ be current estimates for $(\boldsymbol{\Phi}, \boldsymbol{D}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\nu})$. Denoting $\langle \hat{w}_{nk} \rangle \triangleq \mathbb{E}_q[w_{nk}|\boldsymbol{I}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{D}}; \hat{\boldsymbol{\theta}}, \hat{\sigma}^2, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\nu}}]$ and $\langle \hat{w}_{nk}^2 \rangle \triangleq \mathbb{E}_q[w_{nk}^2|\boldsymbol{I}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{D}}; \hat{\boldsymbol{\theta}}, \hat{\sigma}^2, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\nu}}]$, our EM algorithm variant proceeds as follows:

E-Step. Update parameter estimates for the approximating distribution q:

$$\hat{\mu}_{nk} \leftarrow \frac{\left\langle I_n - \sum_{\ell \neq k} \langle \hat{w}_{n\ell} \rangle (\hat{D}_\ell \circ \hat{\Phi}_n^{-1}), \hat{D}_k \circ \hat{\Phi}_n^{-1} \right\rangle - \hat{\sigma}^2 \hat{\theta}_k}{\|\hat{D}_k \circ \hat{\Phi}_n^{-1}\|_2^2}, \tag{6}$$

$$\hat{\nu}_{nk} \leftarrow \frac{\hat{\sigma}^2}{\|\hat{D}_k \circ \hat{\Phi}_n^{-1}\|_2^2}, \tag{7}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard dot product.

Then update $\langle \hat{w}_{nk} \rangle$ and $\langle \hat{w}_{nk}^2 \rangle$ for all n and k:

$$\langle \hat{w}_{nk} \rangle = \hat{\mu}_{nk} + \frac{\sqrt{\hat{\nu}_{nk}} \exp(-\hat{\mu}_{nk}^2/(2\hat{\nu}_{nk}))}{\sqrt{2\pi} Q(-\hat{\mu}_{nk}/\sqrt{\hat{\nu}_{nk}})}, \tag{8}$$

$$\langle \hat{w}_{nk}^2 \rangle = \hat{\nu}_{nk} + \hat{\mu}_{nk}^2 + \frac{\hat{\mu}_{nk} \sqrt{\hat{\nu}_{nk}} \exp(-\hat{\mu}_{nk}^2/(2\hat{\nu}_{nk}))}{\sqrt{2\pi} Q(-\hat{\mu}_{nk}/\sqrt{\hat{\nu}_{nk}})}, \tag{9}$$

[1] Because of approximations we make, the algorithm is strictly speaking not EM or even generalized EM.

where $Q(x) \triangleq \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ is the tail probability of the standard normal distribution.

M-Step. Compute an intermediate deformation estimate $\widetilde{\Phi}_n$ by registering observed image I_n to expected pre-image $\sum_{k=1}^K \langle \hat{w}_{nk} \rangle \hat{D}_k$ using diffeomorphic Demons registration. This step can be computed in parallel across subjects.

After performing all N registrations, enforce that the average deformation is identity. To do this, let $\widetilde{\Phi}_n = \exp(\widetilde{\mathcal{V}}_n)$, where $\widetilde{\mathcal{V}}_n$ is the velocity field of intermediate deformation estimate $\widetilde{\Phi}_n$. Then update $\hat{\Phi}_n$ for all n by computing

$$\hat{\Phi}_n \leftarrow \exp\left(\widetilde{\mathcal{V}}_n - \frac{1}{N}\sum_{m=1}^N \widetilde{\mathcal{V}}_m\right). \tag{10}$$

Next, update parameter estimates for $\boldsymbol{\theta}$ and σ^2:

$$\hat{\theta}_k \leftarrow \frac{1}{\frac{1}{N}\sum_{n=1}^N \langle \hat{w}_{nk} \rangle}, \tag{11}$$

$$\hat{\sigma}^2 \leftarrow \frac{1}{N|\Omega|}\sum_{n=1}^N \left[\left\| I_n - \sum_{k=1}^K \langle \hat{w}_{nk} \rangle (\hat{D}_k \circ \hat{\Phi}_n^{-1}) \right\|_2^2 + \sum_{k=1}^K (\langle \hat{w}_{nk}^2 \rangle - \langle \hat{w}_{nk} \rangle^2) \|\hat{D}_k \circ \hat{\Phi}_n^{-1}\|_2^2 \right]. \tag{12}$$

Finally, update each dictionary element estimate $\hat{D}_k$ while holding the other dictionary elements and parameters constant by numerically minimizing the following energy using projected subgradient descent:

$$E(D_k) = \frac{1}{2\hat{\sigma}^2}\sum_{n=1}^N \sum_{x\in\Omega} |\nabla \hat{\Phi}_n(x)| \left[\left(I_n(\hat{\Phi}_n(x)) - \sum_{\ell=1}^K \langle \hat{w}_{n\ell} \rangle D_\ell(x) \right)^2 + (\langle \hat{w}_{nk}^2 \rangle - \langle \hat{w}_{nk} \rangle^2) D_k^2(x) \right] + \lambda \sum_{x\in\Omega} |D_k(x)| + \frac{\gamma}{2} \sum_{x\in\Omega} \sum_{y\in\mathcal{N}(x)} (D_k(y) - D_k(x))^2, \tag{13}$$

where $|\nabla \hat{\Phi}(x)|$ is the determinant of the Jacobian of $\hat{\Phi}$ with respect to spatial coordinates evaluated at voxel x, and D_k is contiguous and unimodal with $\|D_k\|_2 \leq 1$. At each step of projected subgradient descent, we need to project an input image onto the space of contiguous, unimodal images residing on the ℓ_2 disc. This projection is done by performing a watershed segmentation of a 6mm-FWHM-blurred version of the input image. From the watershed segmentation, we can find voxels corresponding to the largest mode, which we use to mask out the largest mode in the (not blurred) input image. Then we check the ℓ_2 norm of this masked input image and if it's greater than 1, we scale the image to have unit ℓ_2 norm.

Without blurring for the watershed segmentation, two peaks separated by a few voxels that probably correspond to the same parcel would be identified as two separate segments; blurring mitigates the effect of this phenomenon. Importantly, we return to working with the original (not blurred) input image to be projected after we've determined the largest mode. We acknowledge that this is a heuristic method for enforcing contiguity and unimodality and are currently exploring avenues for replacing this heuristic with a more principled prior to force each dictionary element to represent a parcel.

Initialization. We initialize deformation estimates $\hat{\boldsymbol{\Phi}}$ using groupwise functional registration similar to [5] with intensity-equalized diffeomorphic Demons registration [2]. We then apply watershed segmentation with 8mm-FWHM blurring (similar to the unimodal projection) on the resulting functional registration group template to initialize dictionary elements $\hat{\boldsymbol{D}}$ and retain the largest K segments.

Rather than initialize $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\nu}}$, we directly compute guesses for each $\langle \hat{w}_{nk} \rangle$ by solving a least-squares regression problem for each subject:

$$\min_{\langle \hat{w}_{n1} \rangle, \langle \hat{w}_{n2} \rangle, \ldots, \langle \hat{w}_{nK} \rangle} \left\| I_n - \sum_{k=1}^{K} \langle \hat{w}_{nk} \rangle (\hat{D}_k \circ \hat{\Phi}_n^{-1}) \right\|_2, \tag{14}$$

where we set $\langle \hat{w}_{nk} \rangle$ to 0 if its least-squares solution is negative.

Lastly, we compute initial estimates for $\boldsymbol{\theta}$ and σ^2. We use update equation (11) to get an initial estimate for $\boldsymbol{\theta}$. As for σ^2, we use the initial estimate of

$$\hat{\sigma}^2 = \frac{1}{N|\Omega|} \sum_{n=1}^{N} \left\| I_n - \sum_{k=1}^{K} \langle \hat{w}_{nk} \rangle (\hat{D}_k \circ \hat{\Phi}_n^{-1}) \right\|_2^2. \tag{15}$$

3 Experimental Results

We train our model on an fMRI study of 33 subjects reading sentences and pronounceable non-words [3]. First, we apply the standard fMRI general linear model for the sentences vs. non-words contrast. Observed image I_n is defined to be the t-statistic map of subject n thresholded at p-value=0.001. Each image I_n is pre-aligned using an anatomical MRI scan of the same subject. We set hyperparameters $\lambda = 2 \cdot 10^4$ and $\gamma = 10^7$.

During initialization, watershed segmentation yielded $K = 12$ segments that contain at least a pre-specified threshold of 70 voxels each. Fig. 1a shows the spatial support of the final learned dictionary elements on three slices. Fig. 1b illustrates some of the dictionary elements extracted by the algorithm. The dictionary elements correspond to portions of the temporal lobes, the right cerebellum, and the left frontal lobe, regions in the brain previously reported as indicative of lexical and structural language processing [3].

To validate the estimated deformations, we apply the deformation learned by the model to left-out time course data for each subject and perform the standard weighted random effects analysis. We then look at significance values within the

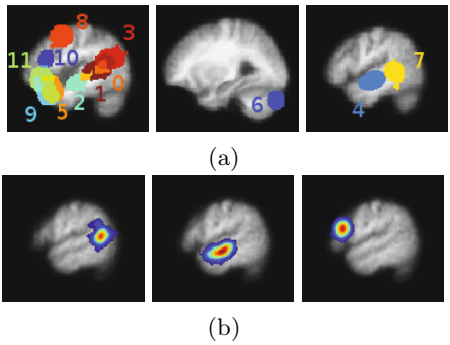

Fig. 1. Estimated dictionary. (a) Three slices of a map showing the spatial support of the extracted dictionary elements. Different colors correspond to distinct dictionary elements where there is some overlap between dictionary elements. From left to right: left frontal lobe and temporal regions, right cerebellum, right temporal lobe. Dictionary element indices correspond to those in Fig. 2. (b) A single slice from three different estimated dictionary volumes. From left to right: left posterior temporal lobe, left anterior temporal lobe, left inferior frontal gyrus.

support of each dictionary element. Importantly, for drawing conclusions on the group-level parcels defined by the estimated dictionary elements, within each parcel, it is the peak and regions around the peak that are of interest rather than the full support of the dictionary element. Thus, to quantify the advantage of our method, within each dictionary element, we compare the top 25% highest significance values for our method versus those of anatomical alignment (Fig. 2). We observe that accounting for functional variability via deformations results in substantially higher peak significance values within the estimated group-level parcels, suggesting better overlap of these functional activation regions across subjects. On average, our method improves the significance of group analysis by roughly 2 orders of magnitude when looking at the top 25% significance values.

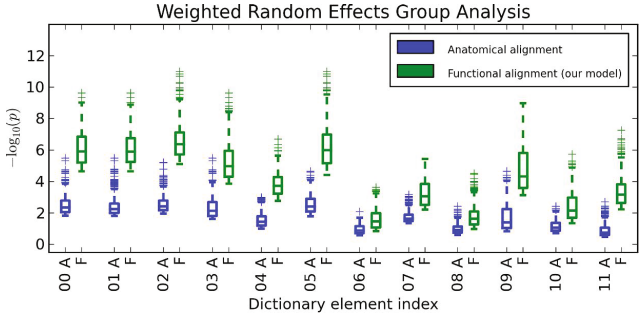

Fig. 2. Box plots of top 25% weighted random effects analysis significance values within dictionary element supports. For each dictionary element, "A" refers to anatomical alignment, and "F" refers to alignment via deformations learned by our model.

Even if we look at the top 50% of significance values in each dictionary element, the results remain similar to those in Fig. 2.

4 Conclusions

We developed a model that accounts for spatial variability of functional regions in the brain via deformations of weighted dictionary elements. Learning model parameters and estimating deformations yields correspondences of functional units in the brain across subjects. We demonstrate our model in a language fMRI study, which contains substantial variability. We plan to validate the detected parcels using data from different fMRI language experiments.

Acknowledgements. This work was funded in part by the NSF IIS/CRCNS 0904625 grant, the NSF CAREER 0642971 grant, the NIH NCRR NAC P41-RR13218 grant, and the NIH NEI 5R01EY13455 grant. George H. Chen was supported by a National Science Foundation Graduate Research Fellowship, an Irwin Mark Jacobs and Joan Klein Jacobs Presidential Fellowship, and a Siebel Scholarship.

References

1. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society 39(1), 1–38 (1977)
2. Depa, M., Sabuncu, M.R., Holmvang, G., Nezafat, R., Schmidt, E.J., Golland, P.: Robust Atlas-Based Segmentation of Highly Variable Anatomy: Left Atrium Segmentation. In: Camara, O., Pop, M., Rhode, K., Sermesant, M., Smith, N., Young, A. (eds.) STACOM-CESC 2010. LNCS, vol. 6364, pp. 85–94. Springer, Heidelberg (2010)
3. Fedorenko, E., Hsieh, P.J., Nieto-Castañón, A., Whitfield-Gabrieli, S., Kanwisher, N.: New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. Neurophysiology 104, 1177–1194 (2010)
4. Olshausen, B., Field, D.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381, 607–609 (1996)
5. Sabuncu, M.R., Singer, B.D., Conroy, B., Bryan, R.E., Ramadge, P.J., Haxby, J.V.: Function-based intersubject alignment of human cortical anatomy. Cerebral Cortex 20(1), 130–140 (2010)
6. Thirion, B., Pinel, P., Tucholka, A., Roche, A., Ciuciu, P., Mangin, J.F., Poline, J.B.: Structural analysis of fMRI data revisited: improving the sensitivity and reliability of fMRI group studies. IEEE Transactions in Medical Imaging 26(9), 1256–1269 (2007)
7. Varoquaux, G., Gramfort, A., Pedregosa, F., Michel, V., Thirion, B.: Multi-subject Dictionary Learning to Segment an Atlas of Brain Spontaneous Activity. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 562–573. Springer, Heidelberg (2011)
8. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Symmetric Log-Domain Diffeomorphic Registration: A Demons-Based Approach. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 754–761. Springer, Heidelberg (2008)
9. Xu, L., Johnson, T.D., Nichols, T.E., Nee, D.E.: Modeling inter-subject variability in fMRI activation location: a bayesian hierarchical spatial model. Biometrics 65(4), 1041–1051 (2009)