Large Deformation Image Classification Using Generalized Locality-Constrained Linear Coding

Pei Zhang[1], Chong-Yaw Wee[1], Marc Niethammer[2], Dinggang Shen[1], and Pew-Thian Yap[1]* for the Alzheimer's Disease Neuroimaging Initiative**

[1] Department of Radiology,
[2] Department of Computer Science,
Biomedical Research Imaging Center (BRIC),
The University of North Carolina at Chapel Hill, USA
peizhang@email.unc.edu, mn@cs.unc.edu,
{chongyaw_wee,ptyap,dgshen}@med.unc.edu

Abstract. Magnetic resonance (MR) imaging has been demonstrated to be very useful for clinical diagnosis of Alzheimer's disease (AD). A common approach to using MR images for AD detection is to spatially normalize the images by non-rigid image registration, and then perform statistical analysis on the resulting deformation fields. Due to the high nonlinearity of the deformation field, recent studies suggest to use initial momentum instead as it lies in a linear space and fully encodes the deformation field. In this paper we explore the use of initial momentum for image classification by focusing on the problem of AD detection. Experiments on the public ADNI dataset show that the initial momentum, together with a simple sparse coding technique—locality-constrained linear coding (LLC)—can achieve a classification accuracy that is comparable to or even better than the state of the art. We also show that the performance of LLC can be greatly improved by introducing proper weights to the codebook.

1 Introduction

Alzheimer's disease (AD) is a brain disorder that causes progressive memory loss and intellectual disabilities. Treatment can be most efficacious if AD can be detected as early as possible. While clinical diagnostic criteria are still routine practice for AD detection, many recent studies have shown that brain imaging, e.g., magnetic resonance imaging (MRI), can greatly facilitate clinical diagnosis.

To use MR images for AD detection, a common approach is to spatially align all the images to a common template and perform statistical analysis on the resulting deformation fields. Examples are statistical deformation models [2] and deformation-based morphometry [7]. However, the deformation field often lies in a highly nonlinear space, which is harder to work with compared with linear space. As such, recent advances in large deformation diffeomorphic metric mapping (LDDMM) methods [5,11,14] suggest to use initial momentum since it lies in a linear space. Moreover, when convolved with an appropriate kernel, the initial momentum generates a series of velocity fields that can be used to obtain the deformation field. As such, the initial momentum uniquely encodes and parameterizes the deformation field.

In this paper we explore the use of initial momentum for image classification, particularly for distinguishing AD or mild cognitive impairment (MCI) patients from healthy controls (HCs). Our work is closely related to [11] and [14]. In [11] principal component analysis (PCA) is applied to a set of initial momenta, and the resulting eigenvalues are used to train a logistic regression model to detect hippocampal shape abnormality in AD. Instead of PCA, Yang et al. [14] used locally linear embedding (LLE) [6] together with initial momenta to represent hippocampal shape in a low-dimensional space.

Our work differs from the above two methods mainly in two aspects. First, we use a recent sparse coding (SC) technique, locality-constrained linear coding (LLC) [10], for classification. Second, our goal is different. We explore the use of initial momentum for image classification, while the above works use it for shape classification [11] and low-dimensional shape embedding [14].

Our main contributions include (1) we propose a generalized LLC that uses a weighted codebook, leading to improved performance; (2) we describe how to use initial momentum to approximate geodesic distance to speed up image coding in the generalized LLC framework.

2   Methodology

Suppose we have a group of training images from multiple classes, a test image and a predefined template image. The problem is to tell which class the test image belongs to. We first align the template to the test image and each training image to obtain a set of initial momenta. This is achieved by the LDDMM algorithm [16], which will be briefly reviewed in Sect. 2.1. Then, for the test image and each training image, their geodesic distance can be approximated using the associated initial momenta as described in Sect. 2.2. Based on the geodesic distance, a subset of training images can be selected as the basis vectors (or codebook). By feeding the codebook to the generalized LLC introduced in Sect. 2.3, we can then classify the test image. Details are given below.

2.1   The LDDMM Algorithm

Let I_0 be the source image and I_1 be the target image. We would like to minimize

$$E(\boldsymbol{v}_t) = \frac{1}{2}\int_0^1 ||\boldsymbol{v}_t||_V^2 \, \mathrm{d}t + \frac{1}{\sigma^2}\,||I_0 \circ \phi_{1,0} - I_1||_{\ell^2}^2\,,$$
$$\text{s.t. } \dot{\phi}_{t,0} + (D\phi_{t,0})\boldsymbol{v}_t = \mathbf{0}, \ \ \phi_{0,0} = \mathtt{id},$$

(1)

where $\boldsymbol{v}_t$ is a time-dependent velocity field to be solved, $\sigma > 0$ is a regularization constant, $\phi_{s,t}$ is a map induced by $\boldsymbol{v}_t$, mapping a voxel from its position at time s to its position at time t, and $\mathtt{id}$ is an identity map. $||\boldsymbol{v}_t||_V^2 = \langle L^\dagger L\boldsymbol{v}_t, \boldsymbol{v}_t \rangle_{\ell^2}$, where L is a proper differential operator. Usually a smoothing kernel $K = (L^\dagger L)^{-1}$ is defined instead of L. $\dot{\phi}_{t,0} = \partial\phi_{t,0}/\partial t$ and $D\cdot$ is the Jacobian operator.

To minimize (1) we use the method of Lagrange multipliers to convert (1) to an unconstrained energy functional, compute the functional variation w.r.t $\boldsymbol{v}_t$, $\phi_{t,0}$ as well as the Lagrange multipliers, and obtain a set of optimality conditions by setting the variation to zero (see [16] for details). Based on the optimality conditions and conservation of momentum (see [5] for details), we can prove that the gradient of (1) w.r.t $\boldsymbol{v}_t$ is equivalent to the gradient w.r.t the initial momentum $\boldsymbol{\lambda}_0$

$$\nabla_{\boldsymbol{\lambda}_0} E = \boldsymbol{\lambda}_0 - |D\phi_{0,1}|\boldsymbol{\lambda}_1 \circ \phi_{0,1}, \tag{2}$$

where $\boldsymbol{\lambda}_1$ is the final adjoint. The proof is beyond the scope of this paper and is thus omitted.

Equation (2) implies that the gradient descent can be directly performed on $\boldsymbol{\lambda}_0$ by pulling $\boldsymbol{\lambda}_1$ back to $t = 0$. The pullback can be achieved by computing a forward map (from $t = 0$ to $t = 1$) on the fly during a backward integration. Specifically, we compute the momentum at time t by $\boldsymbol{\lambda}_t = |D\phi_{t,0}|\boldsymbol{\lambda}_0 \circ \phi_{t,0}$, and then the velocity field at time t by convolving $\boldsymbol{\lambda}_t$ with the kernel K (i.e. $\boldsymbol{v}_t = K\boldsymbol{\lambda}_t$). In this way we can obtain a series of velocity fields $\{\boldsymbol{v}_{t_0}, \boldsymbol{v}_{t_1}, \ldots, \boldsymbol{v}_{t_{N-1}}\}$ at time $\{t_0, t_1, \ldots, t_{N-1}\}$ (suppose we have N uniformly spaced time points). According to [1], the forward map $\phi_{0,1}$ and backward map $\phi_{1,0}$ can be then computed by concatenating a set of small deformation fields, i.e.

$$\begin{aligned}
\phi_{0,1} &= (\mathtt{id} + \frac{1}{N}\boldsymbol{v}_{t_{N-1}}) \circ (\mathtt{id} + \frac{1}{N}\boldsymbol{v}_{t_{N-2}}) \circ \cdots \circ (\mathtt{id} + \frac{1}{N}\boldsymbol{v}_{t_0}), \\
\phi_{1,0} &= (\mathtt{id} - \frac{1}{N}\boldsymbol{v}_{t_0}) \circ (\mathtt{id} - \frac{1}{N}\boldsymbol{v}_1) \circ \cdots \circ (\mathtt{id} - \frac{1}{N}\boldsymbol{v}_{t_{N-1}}).
\end{aligned} \tag{3}$$

Then we can pull $\boldsymbol{\lambda}_1$ back to $t = 0$ by $|D\phi_{0,1}|\boldsymbol{\lambda}_1 \circ \phi_{0,1}$ to update the gradient. In this work we use line search for gradient descent. Note that our formulation avoids the computational complexity of a full adjoint compared with [9].

The above process also reveals that $\boldsymbol{\lambda}_0$ parameterizes the map by $\boldsymbol{\lambda}_t = |D\phi_{t,0}|\boldsymbol{\lambda}_0 \circ \phi_{t,0}$, from which a set of velocity fields can be obtained at each time point. Integrating all the velocity fields over the whole time interval leads to the final map $\phi_{1,0}$. The length ρ of the optimal trajectory from $\phi_{0,0}$ to $\phi_{1,0}$ can be computed as $\rho = \langle \boldsymbol{\lambda}_0, K\boldsymbol{\lambda}_0 \rangle_{\ell^2}$, and represents the geodesic distance between I_0 and I_1 [5].

2.2 Approximate Geodesic Distance

The LLC emphasizes the use of a small set of training images that are most similar to the test image for coding (see Sect. 2.3). This requires us to compute the similarity (or distance) between the test image and each training image, leading to the complexity of $O(M)$, where M is the number of training images. Obviously, this will be very time-consuming if M is large. Below we describe an approximation to the geodesic distance ρ, which will reduce the complexity to $O(1)$.

Let I_T be a template image. We register it to two images I_i and I_j. Ideally, we have $I_T = I_i \circ \phi_{0,1}^i = I_j \circ \phi_{0,1}^j$. Hence, I_j can be represented by I_i using $I_j = I_i \circ \phi_{0,1}^i \circ \phi_{1,0}^j$. By replacing $\phi_{0,1}^i$ and $\phi_{1,0}^j$ with a set of small deformation fields using (3) we have

$$I_j = I_i \circ (\mathtt{id} + \frac{1}{N} \boldsymbol{v}_{t_{N-1}}^i) \circ \cdots \circ (\mathtt{id} + \frac{1}{N} \boldsymbol{v}_{t_0}^i) \circ (\mathtt{id} - \frac{1}{N} \boldsymbol{v}_{t_0}^j) \circ \cdots \circ (\mathtt{id} - \frac{1}{N} \boldsymbol{v}_{t_{N-1}}^j).$$

By expanding the above equation and discarding the higher order terms, we can obtain a first-order approximation to the true map between I_i and I_j by

$$I_j \approx I_i \circ \left(\mathtt{id} + \frac{1}{N}(\boldsymbol{v}_{t_0}^i - \boldsymbol{v}_{t_0}^j) + \frac{1}{N}(\boldsymbol{v}_{t_1}^i - \boldsymbol{v}_{t_1}^j) + \cdots + \frac{1}{N}(\boldsymbol{v}_{t_{N-1}}^i - \boldsymbol{v}_{t_{N-1}}^j) \right).$$

This indicates that we can approximate the geodesic distance between I_i and I_j by

$$\rho \approx \left\langle \boldsymbol{\lambda}_0^i - \boldsymbol{\lambda}_0^j, K(\boldsymbol{\lambda}_0^i - \boldsymbol{\lambda}_0^j) \right\rangle_{\ell^2}. \tag{4}$$

Although a similar result can be found in [14], here we derive from a different perspective. To estimate the distance between the test image and each training image, we only need to register the template image to the test image once and use (4) for approximation (given that we have already registered the template to all the training images).

2.3 Generalized Locality-Constrained Linear Coding

The LLC [10] is a kind of sparse coding (SC) techniques [4,13], where a codebook and a set of weights are used to represent a given data. The weights are sparse in the sense that most are zeros. Compared with other SC approaches, the LLC emphasizes locality, that is, a small set of basis vectors close to the given data instead of the whole codebook should be used for coding. It has been shown that this locality constraint can give better reconstruction and can help avoid similar data to be encoded by different basis vectors.

Let $\boldsymbol{a} \in \mathbb{R}^n$ be the data point to be encoded. Given a set of training data points, the first step of the LLC is to compute the distance between $\boldsymbol{a}$ and each training data point, and choose the k nearest neighbors of $\boldsymbol{a}$ based on the distance in the training set as the codebook. We use $\boldsymbol{B} = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_k\} \in \mathbb{R}^{n \times k}$ to represent the selected codebook. In this work the data point is simply the intensity values across the whole image and the set of initial momenta is used to compute distance as described in Sect. 2.2.

Once we have $\boldsymbol{B}$, we can then code $\boldsymbol{a}$ by solving the following equation

$$\arg \min_{\boldsymbol{c}} ||\boldsymbol{a} - \boldsymbol{B}\boldsymbol{c}||^2 + \alpha ||\boldsymbol{d} \odot \boldsymbol{c}||^2, \ \text{s.t.} \ \mathbf{1}^{\mathrm{T}} \boldsymbol{c} = 1, \tag{5}$$

where $\boldsymbol{c} \in \mathbb{R}^k$ is the weight vector, $\odot$ stands for element-wise multiplication and $\boldsymbol{d} \in \mathbb{R}^k$ represents the distance between $\boldsymbol{a}$ and each basis vector in $\boldsymbol{B}$. Equation (5) essentially estimates a set of weights such that they can be used to best reconstruct $\boldsymbol{a}$ together with the codebook under the locality constraint. However, there are a large number of confounding elements that will adversely affect coding precision. To penalize those confounding elements, we instead use the following equation for coding

$$\arg \min_{\boldsymbol{c}} ||\tilde{\boldsymbol{a}} - \tilde{\boldsymbol{B}}\boldsymbol{c}||^2 + \alpha ||\boldsymbol{d} \odot \boldsymbol{c}||^2, \ \text{s.t.} \ \mathbf{1}^{\mathrm{T}} \boldsymbol{c} = 1, \tag{6}$$

where $\tilde{a} = z \odot a$ and $\tilde{B} = \{w_1 \odot b_1, w_2 \odot b_2, \ldots, w_k \odot b_k\}$. z and $\{w_k\}$ are weighting vectors for penalization. The analytical solution to (6) is given by

$$\tilde{c} = (C + \alpha \text{diag}^2(d)) \backslash 1, \tag{7}$$

$$c = \tilde{c}/1^{\mathrm{T}} c, \tag{8}$$

where $C = (\tilde{B}^{\mathrm{T}} - 1\tilde{a}^{\mathrm{T}})(\tilde{B}^{\mathrm{T}} - 1\tilde{a}^{\mathrm{T}})^{\mathrm{T}}$. In this work z and $\{w_k\}$ are pre-computed using all the data points for efficiency (see Sect. 3 for details).

Once we have the weights c, we use the method described in [12] for classification, i.e. $\arg\min_i ||\tilde{a} - \tilde{B}\hat{c}_i||_{\ell^2}$, where $\hat{c}_i \in \mathbb{R}^k$ is a vector whose nonzero elements are the elements in c associated with class i. As each element of c is associated with a basis b_k, which is, in turn, associated with a class, $\hat{c}_i$ can be obtained by identifying the class of b_k.

3 Experiments

We demonstrate the efficacy of our method using the ADNI database by comparing with one of the state-of-the-art approaches [15]. We used the same dataset as [15] to test our method. The dataset contains 51 AD patients, 99 MCI patients and 52 HCs (see [15] for detailed subject information and the list of subject IDs). We only used the MR images associated with those subjects for our experiment.

We preprocessed the dataset using the FreeSurfer [3]. The procedure includes non-uniform intensity normalization, Talairach transform computation, intensity normalization and skull stripping.

For each image in the dataset we aligned it to a template image using affine registration. Here we used the AAL atlas [8] as the template. Then we registered the template to each of the aligned images using the LDDMM algorithm described in Sect. 2.1. The template image and an example of the initial momentum are shown in Fig. 1a-b. We then used the resulting deformation fields to further warp each affinely aligned image (Fig. 1c) onto the template. Examples of the resulting images are shown in Fig. 1d.

To evaluate the performance of our algorithm we grouped the warped images from the AD patients and the HCs, and randomly split the image pool into 10 even subsets. Note that one subset contains more images than the others since the numbers of images of the AD patients and the HCs are unequal. We evaluated our algorithm using 10-fold cross-validation, which was repeated 10 times with 10 random splits. Similar to [15], we computed the mean of classification accuracy, sensitivity and specificity as performance indicators. We repeated the above process by using the set of warped images grouped from the MCI patients and the HCs.

We first compared the original LLC [10] with the generalized LLC introduced in Sect. 2.3. In all the experiments throughout this paper we set $k = 10$ and $\alpha = 0.001$. We find that smaller k and larger α failed to give satisfactory results. For the generalized LLC, we pre-computed z and $\{w_k\}$ using all the warped images from the two classes (e.g., AD and HC or MCI and HC). This was achieved by performing Welch's (two-tailed) t-test between voxels at corresponding positions in the images. We used hard assignment, that is, all voxels with p-value below a threshold (0.05) are ones, otherwise zeros. The resulting image was converted into a column vector and used as w_k.

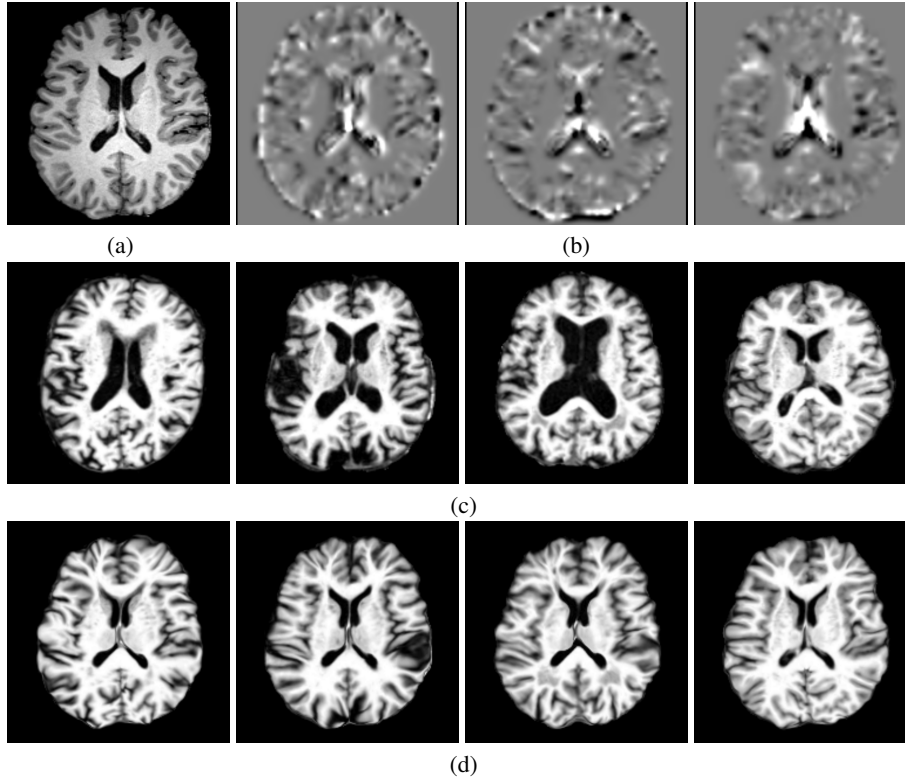(a)   (b)

(c)

(d)

Fig. 1. (a) The template image. (b) An example of the initial momentum in X, Y and Z directions (from left to right). (c) Examples of affinely registered images. (d) Examples of the registered images using the LDDMM algorithm. Note that the initial momentum has been normalized for a better view.

In this work we used the same w_k to weight each basis vector in the selected codebook and set $z = w_k$. Hence, the test images were also weighted using w_k. Although more sophisticated methods can be used to compute z and $\{w_k\}$, below we show that this simple technique is good enough to lead to promising results. A similar t-test was also performed on the set of initial momenta in each direction. The resulting vector image was used to weight each initial momentum, and the geodesic distance was estimated as described in Sect. 2.2 using the weighted momenta, which are found to give a better estimate of distance than the original momenta. We ran our algorithm using the two kinds of LLC respectively. The results shown in Fig. 2 clearly indicate that the generalized LLC significantly outperforms the original one.

We further compared our approach with [15], where linear support vector machines are applied to multi-modal data for classification. We summarize the comparison in Fig. 3. We can see that our method significantly outperforms [15] when only MR images are used. Compared with the results obtained by [15] on multi-modal data, our algorithm
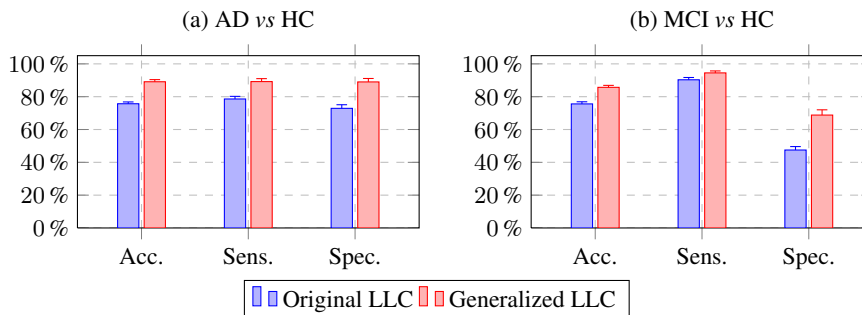
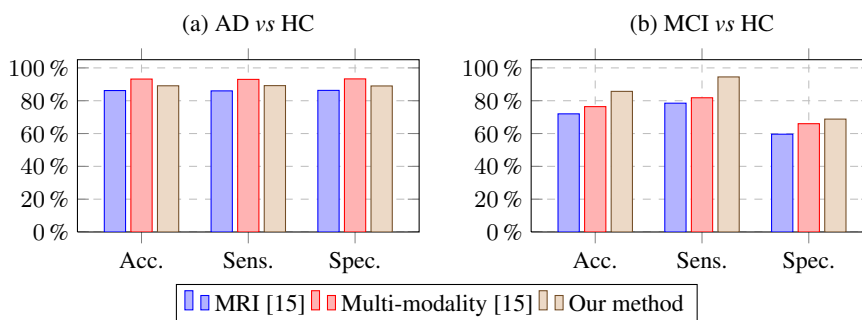**Fig. 2.** Comparison of the original LLC and the generalized LLC.

Fig. 3. Comparison of our method and the work in [15].

yields comparable classification accuracy when distinguishing AD patients from HCs, and significantly better result when distinguishing MCI patients from HCs.

4 Discussion and Conclusions

We have described an image classification method by virtue of the initial momentum. This is achieved by using a generalized LLC algorithm together with an approximate geodesic distance between images. Experimental results show that the generalized LLC algorithm significantly outperforms the original one [10]. This demonstrates that the performance can be greatly improved by focusing on the discriminative voxels during coding. We also show that our algorithm can achieve comparable or even better classification accuracy rate when compared with one of the state-of-the-art methods.

We implemented the system in C++ using the Insight Segmentation and Registration Toolkit[1] (ITK). The typical running time of our LDDMM algorithm is 8∼10 minutes. The timing is based on an iMac with an Intel® Core™ i5 processor (3.1 GHz).

In the future we will explore how to extract and select various features from the initial momentum to further improve classification accuracy. We also plan to investi-

---

[1] http://www.itk.org/

gate more sophisticated methods to learn the weighting matrix for the generalized LLC algorithm.

References

1. Ashburner, J., Friston, K.J.: Diffeomorphic registration using geodesic shooting and Gauss-Newton optimisation. NeuroImage 55(3), 954–967 (2011)
2. Caban, J., Rheingans, P.: Relational statistical deformation models for morphological image analysis and classification. In: Proceedings of International Symposium on Biomedical Imaging. pp. 1333–1336 (2010)
3. Fischl, B.: Freesurfer. NeuroImage 62(2), 774–781 (2012)
4. Lee, H., Battle, A., Raina, R., Ng, A.: Efficient sparse coding algorithms. In: Proceedings of Advances in Neural Information Processing Systems. pp. 801–808 (2007)
5. Miller, M.I., Trouvé, A., Younes, L.: Geodesic shooting for computational anatomy. Journal of Mathematical Imaging and Vision 24(2), 209–228 (2006)
6. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500), 2323–2326 (2000)
7. Teipel, S.J., Born, C., Ewers, M., Bokde, A.L., Reiser, M.F., Mller, H.J., Hampel, H.: Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. NeuroImage 38(1), 13–24 (2007)
8. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15(1), 273–289 (2002)
9. Vialard, F.X., Risser, L., Rueckert, D., Cotter, C.: Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. International Journal of Computer Vision 97(2), 229–241 (2012)
10. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 3360–3367 (2010)
11. Wang, L., Beg, F., Ratnanather, T., Ceritoglu, C., Younes, L., Morris, J.C., Csernansky, J.G., Miller, M.I.: Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the Alzheimer type. IEEE Transactions on Medical Imaging 26(4), 462–470 (2007)
12. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(2), 210–227 (2009)
13. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 1794–1801 (2009)
14. Yang, X., Goh, A., Qiu, A.: Locally linear diffeomorphic metric embedding (LLDME) for surface-based anatomical shape modeling. NeuroImage 56(1), 149–161 (2011)
15. Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D.: Multimodal classification of Alzheimer's disease and mild cognitive impairment. NeuroImage 55(3), 856–867 (2011)
16. Zhang, P., Niethammer, M., Shen, D., Yap, P.T.: Large deformation diffeomorphic registration of diffusion-weighted images with explicit orientation optimization. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. Springer, Heidelberg (2013 To appear)