SEMI-COUPLED DICTIONARY LEARNING FOR DEFORMATION PREDICTION

Tian Cao[1]*, Nikhil Singh*[1]*, Vladimir Jojic*[1]*, Marc Niethammer*[1,2]

[1]University of North Carolina at Chapel Hill, NC, USA
[2]Biomedical Research Imaging Center, UNC-Chapel Hill, NC, USA

ABSTRACT

We propose a coupled dictionary learning method to predict deformation fields based on image appearance. Rather than estimating deformations by standard image registration methods, we investigate how to obtain a basis of the space of deformations. In particular, we explore how image appearance differences with respect to a common atlas image can be used to predict deformations represented by such a basis. We use a coupled dictionary learning method to jointly learn a basis for image appearance differences and their related deformations. Our proposed method is based on local image patches. We evaluate our method on synthetically generated datasets as well as on a structural magnetic resonance brain imaging (MRI) dataset. Our method results in an improved prediction accuracy while reducing the search space compared to nearest neighbor search and demonstrates that learning a deformation basis is feasible.

Index Terms— dictionary learning, deformation prediction

1. INTRODUCTION

Image registration is a critical medical image analysis task to establish spatial correspondences between images. Standard image registration approaches are based on the numerical solution of optimization problems. However, recent work has focused on *learning* registration maps using example deformations, which can then be used to predict deformations instead of optimizing with respect to them. For instance, using machine learning methods, Chou et al. [1] and Wang et al. [2] propose prediction based models to estimate deformations based on image appearances. The deformation prediction problem is challenging as it is difficult to model the relationship between image appearances and deformations. For example, the relationship between appearances and local deformations could be highly nonlinear. Moreover, little work has been done to explore the distribution of image appearance and deformations jointly. Our proposed coupled dictionary learning method addresses this issue: it considers the distribution of both spaces together to learn bases for both image appearance as well as the associated deformations. This is important (1) to establish if such bases can be learned, (2) to establish if appearance differences can predict deformation differences, and (3) to understand how appearance differences relate to deformations.

In [1], the authors propose to learn a global correlation between image appearance and the deformation or the parameters for deformation. Image intensity differences or image level features are used to learn regression models for the corresponding deformation parameters. For a new test image, the regression models are applied to predict the deformation field. In [2], the authors predict the deformations on some key points in a test image from a sparse linear combination of deformations from training images based on the similarity of appearances. A dense deformation field is then interpolated by thin-plate splines.

This paper presents a coupled dictionary learning method to predict deformations based on a sparse representation model on image patches. Coupled dictionary learning methods learn a joint dictionary on two different spaces to establish correspondence of dictionary elements. Dictionary learning plays a key role in many applications using sparse models [4, 5]. In [4], a coupled dictionary learning is performed for image super-resolution. A multi-modal dictionary, a special case of coupled dictionary, is learned from correlative microscopy images and applied to multi-modal registration in [5]. While in [5], the two parts of the coupled dictionary are assumed equivalent to each other, Wang et al. [4] establish an explicit mapping between the two dictionaries. We focus on coupled dictionary learning for *deformation prediction from appearance differences* in this paper.

Our main contributions include:

- a dictionary based framework for deformation prediction using appearance;
- a method for learning a dictionary that accounts for disparity of appearance and transformation spaces;
- an illustration of the applicability to different types of deformations.

The paper is organized as follows: Sec. 2 describes the coupled dictionary learning method. Sec. 3 introduces momenta for deformation parametrization [1]. Sec. 4 applies our model

[1]We also tested our method on B-spline transformation, but omitted the results due to page constraints.

to both synthetic and real data. The paper concludes with a summary of results and an outlook on future work in Sec. 5.

2. METHOD

Image registration estimates a transformation between a source image and a target image. Let $\phi(x)$ denote a transformation that moves a pixel at a location x in the image to another location, $y = \phi(x)$. Suppose the source (moving) image, S, is an atlas and the target (static) image, T, corresponds to an image of a subject. The goal is to estimate the deformations $\phi(\cdot)$ for the entire atlas image that transforms it to match the subject's image in the sense that, the appearance at a pixel in the deformed source image, $\tilde{S}(y) = S(\phi^{-1}(y))$ best matches with the target image, $T(y)$, at pixel y.

There are several ways to parametrize the deformation function, ϕ. We will discuss the parametrization of deformation with initial momenta in Sec. 3. In our framework, we first extract patches from both training images and the corresponding deformations, and train a coupled dictionary jointly on the patches. Here, the training image patches are the patch-wise intensity differences between training subject images and the atlas image in the atlas space. Given a test image, we reconstruct its intensity difference with respect to the atlas image patch by patch with the dictionary corresponding to the image appearance at the same time predicting the deformation with the dictionary corresponding to the deformation. The overall predicted deformation is reconstructed from local patches. We average reconstruction results from overlapping patches to minimize artifacts.

2.1. Coupled Dictionary Learning based on Sparse Representation Model

Learning a dictionary D under a sparse representation model corresponds to solving the optimization problem,

$$\{\hat{D}, \hat{\alpha}\} = \operatorname*{argmin}_{D,\alpha} \sum_{i=1}^{N} \frac{1}{2}\|x_i - D\alpha_i\|_2^2 + \lambda\|\alpha_i\|_1, \quad (1)$$

where $x_i \in \mathbb{R}^p$ is the training data, N is the number of training samples. The ℓ_1 regularity term induces sparsity in the coefficients α_i for the dictionary atoms (columns of D) to approximate x_i. To avoid D being arbitrarily large, each column of D is normalized to have ℓ_2 norm less than or equal to one, i.e. $d_k^T d_k \leq 1$, for $k = 1, ..., p$, and $D = \{d_1, d_2, ..., d_q\} \in \mathbb{R}^{p\times q}$.

Similarly, a *coupled* dictionary learning (CDL) can be formulated as

$$\{\hat{\tilde{D}}, \hat{\alpha}\} = \operatorname*{argmin}_{\tilde{D},\alpha} \sum_{i=1}^{N} \frac{1}{2}\|\tilde{x}_i - \tilde{D}\alpha_i\|_2^2 + \lambda\|\alpha_i\|_1, \quad (2)$$

where $\tilde{D} = [D^1, D^2]^T$ stacks two different dictionaries and $\tilde{x}_i = [x_i^1, x_i^2]^T$ is the corresponding stacked training data from two different spaces.

The coupled dictionary is learned on the data from two different spaces, here, image appearance space and deformation parameter space. Using only one set of coefficients imposes the strong assumption that the coefficients of the representation of the two spaces are equal. To relax this assumption, a semi-coupled dictionary learning (SCDL) method was proposed in [4],

$$\begin{aligned}\{\hat{D}^1, \hat{D}^2, \hat{\alpha}^1, \hat{\alpha}^2, \hat{W}\} = &\operatorname*{argmin}_{D^1,D^2,\alpha^1,\alpha^2,W} \\ &\sum_{i=1}^{N} \frac{1}{2}\|x_i^1 - D^1\alpha_i^1\|_2^2 + \frac{1}{2}\|x_i^2 - D^2\alpha_i^2\|_2^2 + \\ &\lambda_1\|\alpha_i^1\|_1 + \lambda_2\|\alpha_i^2\|_1 + \gamma_1\|\alpha_i^1 - W\alpha_i^2\|_2^2 + \gamma_2\|W\|_F^2,\end{aligned} \quad (3)$$

where λ_1, λ_2, γ_1, γ_2 are regularization parameters. Distinct from CDL, W is a matrix defining a mapping between the coefficients in the two spaces. Unlike for CDL the columns of the dictionaries D^1 and D^2 are normalized separately. As we will show in Sec. 4 such a separate normalization is beneficial to jointly compute a basis for appearance differences and deformations. Eq. (3) is not convex with respect to D^1, D^2, α^1, α^2, W jointly, however, it is convex with respect to each of them when others are fixed.

2.2. Deformation Estimation

After obtaining D^1, D^2 and the linear mapping W from training data x_i^1 and x_i^2, given a difference image $I = T - S$, where S is an input source image, and T is the common target/atlas image, similar to Eq. (3), we solve the following

$$\{\alpha_i^1\} = \operatorname*{argmin}_{\alpha_i^1} \frac{1}{2}\|I_i^1 - D^1\alpha_i^1\|_2^2 + \lambda_1\|\alpha_i^1\|_1, \quad (4)$$

where I_i is a patch of I [2]. Eq. (4) is a sparse coding problem and can be solved using SPAMS [6]. The corresponding deformation parameters p_i of I_i can be estimated as,

$$p_i = D^2 W \alpha_i^1.$$

Here p_i are the parameters defining the deformation ϕ_i of the ith patch. After estimating ϕ_i from p_i for all the patches, we can determine the overall ϕ.

3. PARAMETRIZATION FOR DEFORMATION WITH INITIAL MOMENTUM

Given a source image I and a target image T, the deformation ϕ establishes a mapping between the coordinates of I and T, i.e. $I(\phi(x)) = T(x)$ and $T(\phi^{-1}(y)) = I(y)$, where x and y are vectors of the coordinates in T and I respectively.

[2] We reshape the image patch to a vector.

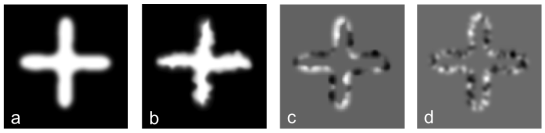

Fig. 1: Local deformation experiment: Illustration of synthetic (a) atlas, (b) deformed training, and (c) difference image between atlas and deformed images; (d) corresponding initial scalar momentum. (Intensities in (c) and (d) scaled for better visualization.)

Many different parametrizations exist [7]. Here, we use initial momenta to parametrize deformations within the large deformation diffeomorphic metric mapping (LDDMM) framework [8, 9] which are elements of a vector space and hence a convenient parametrization in the context of the averaging performed in sparse coding and dictionary learning. Given a source I_0 and target image I_1 LDDMM solves

$$\min_{v_t} \frac{1}{2} \int_0^1 \|v_t\|_V^2 \mathrm{d}t + \frac{1}{\sigma^2} \|I_0 \circ \phi_{1,0} - I_1\|_2^2, \quad \text{s.t. } \dot{\phi}_{t,0} + (D\phi_{t,0})v_t = 0, \phi_{0,0} = \mathtt{id}, \tag{5}$$

where v_t is the velocity vector field at time t, $t \in [0,1]$, $\|v_t\|_V^2 =< L^\dagger L v_t, v_t >_2$, L is a differential operator, $\phi_{s,t}$ defines a mapping for a voxel from its position at time s to its position at time t. $\dot{\phi}_{t,0} = \partial \phi_{t,0} / \partial t$ and D is the Jacobian operator. The initial momentum, m, is defined by the linear mapping, $m = (L^\dagger L) v_0$. It completely parametrizes the geodesic connecting the source and target images. For a more in-depth discussion about LDDMM image registration and momentum based parametrization, refer to [10].

4. EXPERIMENTAL VALIDATION

We present two experiments to validate our proposed method: a synthetic experiment illustrating the behavior for diffeomorphic transformations (Sec. 4.1), and an experiment on real brain data (Sec. 4.2).

4.1. Experiment on Synthetic Data with Local Deformations

In this experiment, we tested our methods on random local deformations. We used a smoothed cross image (Fig. 1) as atlas. The training data were generated by applying random deformations to the atlas. The deformations were parameterized by scalar initial momentum [9]. The random initial momenta were generated by random sampling from a Gaussian distribution with $\sigma = 3$ (corresponding to a maximum displacement of about 10 pixels).

We trained coupled dictionaries based on the intensity differences between atlas and deformed training images and their initial momenta. The image size is 128×128, and we extracted 15×15 image patches from 200 difference images to train the coupled dictionaries. The image patches are extracted with 1 pixel stride, obtaining overlapping patches. During the testing, we first reconstructed each local patch, then reconstructed the whole image by averaging the overlapped patches [5]. We compared the SCDL, CDL and NN methods on deformation prediction for 50 test images. Tab. 1 shows the results of deformation prediction by comparing the transformed test images with the predicted deformations to the atlas images. Both SCDL and CDL methods show better performance compared with NN. A dictionary size of 1000 is sufficient for deformation prediction. Increasing dictionary size to 5000 does not significantly (p-value$_{SCDL} = 0.97$, p-value$_{CDL} = 0.99$) improve performance.

Table 1: Registration results for predicting random initial momenta for synthetic data (Sec. 4.1). Statistics for absolute errors in pixels of predicted deformation to the ground truth. RAW indicates images without registration, NN indicates nearest neighbor search and CDL/SCDL are coupled and semi-coupled dictionary learning methods respectively.

Method	median	mean	min	max	std
RAW	6.064	6.167	3.754	9.557	1.719
NN	3.431	3.662	1.739	6.876	1.911
CDL100	3.315	3.255	0.924	4.843	1.325
CDL500	2.354	2.316	0.532	3.681	1.084
CDL1000	1.819	1.676	0.262	2.739	1.114
CDL5000	1.525	1.570	0.222	2.568	1.181
SCDL100	2.841	2.646	0.851	4.245	1.099
SCDL500	1.805	1.728	0.320	3.027	1.111
SCDL1000	1.333	1.233	0.229	2.082	0.76
SCDL5000	**1.225**	**1.120**	**0.223**	**1.858**	**0.654**

Table 2: Registration results for predicting initial momenta for OASIS dataset (Sec. 4.2). Statistics for absolute errors in pixels of predicted deformation to the ground truth.

Method	median	mean	min	max	std
RAW	7.517	7.495	5.250	9.935	1.186
NN	3.572	3.753	1.221	8.644	1.790
CDL100	3.430	3.455	0.915	6.122	1.060
CDL500	2.426	2.446	1.143	3.681	0.568
CDL1000	1.988	1.867	0.810	3.310	0.660
CDL5000	1.750	1.691	0.802	3.110	1.052
SCDL100	2.954	2.757	1.083	4.232	0.765
SCDL500	1.855	1.831	0.619	3.027	0.751
SCDL1000	1.439	1.548	0.685	2.782	0.629
SCDL5000	**1.402**	**1.455**	**0.614**	**2.529**	**0.547**

4.2. Experiment on Real Data

The previous experiments were based on synthetic data. Here, we test our method on the Open Access Series of Imaging

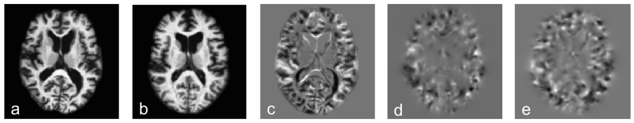

Fig. 2: Illustration of brain (a) atlas image, (b) subject image, (c) difference image between atlas and subject images and (d,e) corresponding initial momentum in x and y direction respectively for experiment on OASISs dataset. (Intensities in (c), (d) and (e) scaled for better visualization.)

Studies (OASIS) Cross-sectional MRI data [11]. We constructed an atlas from 100 subject images with LDDMM to obtain the initial momenta [9]. Fig. 2 shows the atlas image and an example subject image. The coupled dictionary is learned based on intensity difference between atlas and subject images and the corresponding initial momenta. We extracted 15×15 image patches from 128×128 training images. We tested our method on 50 subject images. Tab. 2 shows the results of deformation prediction by comparing the deformed test images with the predicted deformations to the atlas images. The SCDL method shows better performance compared with NN and CDL at the same time reducing the search space compared to the NN method. A dictionary size of 1000 is sufficient for deformation prediction. Increasing dictionary size to 5000 does not significantly (p-value$_{SCDL} = 0.98$, p-value$_{CDL} = 0.98$) improve performance.

5. DISCUSSION AND CONCLUSION

We proposed two coupled dictionary learning methods (CDL and SCDL) for deformation prediction from image appearance differences through a regression model. Both dictionary learning methods are capable of learning a basis relating appearance differences to deformations. In particular, they allow for faithful deformation prediction using moderately sized dictionaries even for high-dimensional appearance and deformation spaces. Our experiments show that prediction performance for CDL and SCDL saturates when moving towards larger dictionaries, indicating that the learning procedure is able to capture a meaningful basis for the observed deformations. Consequentially, both methods generalize better than the NN method when training data is scarce in comparison to the deformation space, which is the case for general deformable registration. SCDL further improves performance over CDL, because it enables flexible coupling between the appearance and the deformation spaces and provides an improved way of learning dictionaries through independent normalization.

Another application of our model is to study how appearance influences the shape of an object. This could be useful where biological processes within an object drive the movement of its boundaries, e.g., when studying how cell-signaling (as imaged through a biosensor) relates to cell shape changes. In this case the learned relation between appearance and deformation will be the objective itself rather than using a predicted deformation to obtain image correspondences.

6. REFERENCES

[1] Chen-Rui Chou, Brandon Frederick, Gig Mageras, Sha Chang, and Stephen Pizer, "2d/3d image registration using regression learning," *CVIU*, vol. 117, no. 9, pp. 1095–1106, 2013.

[2] Qian Wang, Minjeong Kim, Guorong Wu, and Dinggang Shen, "Joint learning of appearance and transformation for predicting brain mr image registration," in *IPMI*, 2013, pp. 499–510.

[3] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T.S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031–1044, 2010.

[4] Shenlong Wang, D Zhang, Yan Liang, and Quan Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *CVPR*, 2012, pp. 2216–2223.

[5] Tian Cao, Christopher Zach, Shannon Modla, Debbie Powell, Kirk Czymmek, and Marc Niethammer, "Multimodal registration for correlative microscopy using image analogies," *Medical image analysis*, 2013.

[6] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009, pp. 689–696.

[7] M. Holden, "A review of geometric transformations for nonrigid body registration," *IEEE Transactions on Medical Imaging*, vol. 27, no. 1, pp. 111–128, 2008.

[8] Michael I Miller, Alain Trouvé, and Laurent Younes, "Geodesic shooting for computational anatomy," *Journal of mathematical imaging and vision*, vol. 24, no. 2, pp. 209–228, 2006.

[9] Nikhil Singh, Jacob Hinkle, Sarang Joshi, and P Thomas Fletcher, "A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction," in *ISBI*, 2013, pp. 1219–1222.

[10] Laurent Younes, *Shapes and Diffeomorphisms*, vol. 171, Springer, 2010.

[11] Daniel S Marcus, Tracy H Wang, Jamie Parker, John G Csernansky, John C Morris, and Randy L Buckner, "Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," *Journal of cognitive neuroscience*, vol. 19, no. 9, pp. 1498–1507, 2007.