Coupled Dictionary Learning for Image Analysis

Tian Cao

Department of Computer Science
UNC-Chapel Hill, USA

Abstract

Modern imaging technologies provide us different ways to visualize various objects ranging from molecules in a cell to the tissue of a human body. Images from different imaging modalities reveal distinct information about these objects. Thus a common problem in image analysis is how to relate different information about the objects. For example, joint analysis of correlative microscopic images requires aligning images from different modalities; estimating the deformation of a subject image with respect to a common reference image, also called an atlas image, based on the appearances of images requires modeling the relationship between deformations and image appearances; exploring the coordination of GTPase activations and cell protrusion of mouse embryonic fibroblasts requires uncovering the spatiotemporal correspondences between GTPase activations and cell movements. These problems are challenging due to the difficulties to model the relationship between the information from different modalities/sources. For example, the correlation between image appearances and image deformations could be highly nonlinear. I propose using a coupled dictionary learning based approach to capture the relationship between the data from two spaces. Coupled dictionary learning is a method to learn a coupled basis for the compact representation of data in two spaces. In this report, I first introduce a coupled dictionary learning based image analogies method. Multi-modal image registration (for example, registration between correlative microscopic images) can be simplified to a mono-modal one by this method. Furthermore, I propose a semi-coupled dictionary learning based framework to estimate the deformations from image appearances. Moreover, I explore the use of coupled dictionary learning method to capture the relationship between GTPase activations and cell protrusions and retractions. Finally, I propose a probabilistic model for robust coupled dictionary learning to address learning a coupled dictionary with non-corresponding data. This method discriminates between corresponding and non-corresponding data thereby resulting in a "clean" coupled dictionary by removing non-corresponding data during the learning process.

1 Introduction

With the development of new imaging technologies, we can visualize various objects to explore information ranging from molecular structures of cell to tissue textures of the human body. Different imaging modalities provide distinct information of the objects. For example, in the context of correlative microscopy, protein locations can be revealed through fluorescence microscopy, while protein structures can be observed through electron microscopy [8].

Many image analysis applications require relating the information from different modalities/sources. For example, joint analysis of correlative microscopic images needs registration of images from different modalities; estimating the deformation of an image with respect to an atlas requires modeling the relationship between deformations and image appearances; investigating the coordination of GTPase activities and cell protrusions of mouse embryonic fibroblasts (MEFs) requires establishing the spatiotemporal correspondences between GTPase activations and cell movements.

The similarity between these applications is that they require us to *model the relationship of data from different spaces.* I propose coupled dictionary learning based approaches to *learn* these relationships from given data.

Correlative microscopy is a methodology combining the functionality of light microscopy with the high resolution of electron microscopy and other microscopy technologies. The first step of most correlative microscopy based applications is to do registration between two or more microscopic images [7]. Image registration is the process of estimating spatial transformations between images (to align them). The registration of correlative microscopic images is challenging: images could carry distinct information to combine, for example, knowledge about protein locations (using fluorescence microscopy) with high-resolution structural data (using electron microscopy).

I propose a coupled dictionary learning based image analogies[1] *method thereby allowing for the reconstruction of a microscopy image in the appearance space of another.* This approach can simplify the multi-modal registration problem of correlative microscopy to a mono-modal one.

A dictionary, i.e., a set of basis signals, is usually learned in a sparse representation model. Such a model represents the signal with sparse combinations of atoms in a learned dictionary. In my approach, I learn a joint dictionary for the appearance of two different modalities with a coupled dictionary learning method. Coupled dictionary learning is a special dictionary learning method that learns a *joint* dictionary on two different spaces to establish correspondence of dictionary atoms [29].

Furthermore, some medical image analysis tasks, such as brain image analysis, require registering subject images to a common atlas (reference) image, i.e, estimating the underlying deformation fields that transform the subject images to the atlas image. Recent work has focused on learning registration maps using example deformations, which can then be used to predict deformations with respect to them [9, 24].

[1] An image synthesis approach which synthesize a “filtered” version of an input image based on the relationship between training images A and A', where A' is a “filtered” version of A [13].

The deformation estimation problem is challenging as (i) it is difficult to model the relationship between image appearances and deformations. For example, the correlation between appearances and deformations could be highly nonlinear; (ii) different deformation representation methods may further complicate the modeling of the relationship.

I propose a general framework to estimate the deformation field for a given subject image with respect to an atlas image based on coupled dictionary learning for image appearance and the corresponding deformation fields. The relationship between appearances and deformations can be captured by the coupled dictionary. This then allows for the estimation of deformations based on the appearances of a test image and the coupled dictionary in a sparse representation model.

Moreover, exploring the spatiotemporal coordination of GTPase activations and cell protrusions is crucial for the understanding of cell dynamics [17, 20]. GTPases (Rac1, RhoA and Cdc42) are a family of enzymes that can bind GTP. GTPase activities can be studied with biosensor imaging. This problem is challenging for the following reasons: (i) a good data representation is required to capture the relationship between GTPase activations and cell protrusions; (ii) establishing the spatiotemporal correspondences between GTPase activations and cell protrusions is difficult as both the shapes and appearances (activations) of cells change during cell movements.

I propose to learn a coupled dictionary on GTPase activities and cell protrusions of MEFs. Common spatiotemporal patterns for the GTPase activation corresponding to the cell edge movement can then be obtained by coupled dictionary learning.

In summary, coupled dictionary learning plays a key role in the previously discussed applications, but presents some challenges: (i) it may fail without sufficient correspondences between the data from different spaces, for example, a low quality image deteriorated by noise in one modality can hardly match a high quality image in another modality. (ii) Usually the data correspondence for two different spaces needs to be obtained before learning the coupled dictionary, for example, pre-registering training images before dictionary learning. Errors can be introduced during the process of establishing these correspondence.

Thus, *I propose a robust dictionary learning method under a probabilistic model.* Instead of directly learning a coupled dictionary from training data, I distinguish between image regions with and without good correspondence in the learning process and update the learned dictionary iteratively.

1.1 Contributions

The main contributions of my work include:

- a sparse representation and a coupled dictionary learning based image analogies method to convert and thereby simplify a multi-modal registration problem to a mono-modal one;

- a general framework for deformation estimation using appearance information based on coupled dictionary learning;

- a framework for relating the spatiotemporal patterns between GTPase activations and cell protrusion of MEFs based on coupled dictionary learning;
- a robust coupled dictionary learning method based on a probabilistic model which discriminates between corresponding and non-corresponding patches automatically.

1.2 Thesis

Learning a coupled basis for the compact representation of two spaces can be achieved by coupled dictionary learning. Such dictionaries can be learned to capture appearance differences of different imaging modalities, dependencies between image appearance and deformation as well as the spatiotemporal patterns for cell signaling and boundary protrusions and retractions. To account for data inconsistencies a robust coupled dictionary can be obtained based on a probabilistic dictionary model.

The remainder of the document is organized as follows: Section 2 introduces some background. Section 3 describes the image analogies based multi-modal registration. The joint dictionary learning method for both appearances and deformations is explained in Section 4. Section 5 presents an application of coupled dictionary learning for modeling GTPase activation and cell protrusion for MEFs. A robust multi-modal dictionary learning method is proposed in section 6. Section 7 discusses some future work. A summary is presented in Section 8.

2 Background

2.1 Related Work

2.1.1 Multi-modal Image Registration for Correlative Microscopy

A possible solution for the registration of correlative microscopy images is to perform landmark-based alignment, which can be greatly simplified by adding fiducial markers [12]. Fiducial markers cannot easily be added to some specimen, hence an alternative image-based method is needed. This can be accomplished in some cases by appropriate image filtering. This filtering is designed to only preserve information which is indicative of the desired transformation, to suppress spurious image information, or to use knowledge about the image formation process to convert an image from one modality to another. E.g., multichannel microscopy images of cells can be registered by registering their cell segmentations [31]. However, such image-based approaches are highly application-specific and difficult to devise for the non-expert.

Standard methods for general multi-modal image registration include (i) applying advanced similarity measures, such as mutual information (MI) [27], (ii) or transforming a multi-modal to a mono-modal registration problems [26]. For example, Wachinger introduced entropy images and Laplacian images which are general structural representations

[23]. The motivation of my proposed method is similar to Wachinger's approach, i.e. transforming the modality of one image to another, but I use image analogies to achieve this goal thereby allowing for the reconstruction of a microscopy image in the appearance space of another by *learning* this transformation.

2.1.2 Deformation Estimation from Appearances

In [9, 15], 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 on the corresponding deformation parameters. For a new test image, the regression models are applied to predict the deformation field. In [21, 24], the authors predict the deformations for a few key points in a test image from a sparse linear combination of deformations from training images based on the similarity of appearances in [21], and then a dense deformation field is interpolated by Thin-Plate Spline model in [24]. These methods are based on finding the correspondences between the key points in the test image and the training images. However, the key points are not always available in images and it is not clear how to find these key points.

2.1.3 Coordination between GTPase Activations and Cell Protrusions

GTPases play key roles in controlling cell dynamics [20, 17]. In [16], the authors propose a framework for tracking cell boundary movements. The level set method is used to reconstruct the continuous boundary movement in image sequences with finite time sampling. Based on this method, the authors propose an approach to examine the spatiotemporal coordination between GTPase activations and cell protrusion in mouse embryonic fibroblasts in [17]. The protrusion and retraction of cell boundary are defined as the paths of boundary points traverses when moving orthogonal to the cell boundary. The location of the cell edge is tracked and compared to the edge velocities with the biosensor signal (image intensities) in regions moving with the leading edge. However, the regions of interest are hand-picked and the movement information in other regions is ignored in this method.

2.2 Sparse Representation Model

Sparse representation is a technique to reconstruct a signal as a combination (usually linear) of a few basis signals from a typically over-complete dictionary [3]. A dictionary is a collection of basis signals. The number of dictionary elements in an over-complete dictionary exceeds the dimension of the signal space (here the dimension of an image patch). Suppose a dictionary D is pre-defined. To sparsely represent a signal x the following optimization problem is solved [10]:

$$\hat{\alpha} = \arg\min_{\alpha} \|\alpha\|_0, \quad \text{s.t.} \quad \|x - D\alpha\|_2 \leq \epsilon, \tag{1}$$

where α is a sparse vector that explains x as a linear combination of columns in dictionary D with error ϵ and $\|\cdot\|_0$ indicates the number of non-zero elements in the vector α. Solving (1) is an NP-hard problem [10]. One possible solution of this problem is based on a relaxation that replaces $\|\cdot\|_0$ by $\|\cdot\|_1$, where $\|\cdot\|_1$ is the ℓ_1 norm of a vector. The equivalent Lagrangian form of (1) is

$$\hat{\alpha} = \arg\min_{\alpha} \lambda\|\alpha\|_1 + \|x - D\alpha\|_2^2, \tag{2}$$

which is a convex optimization problem that can be solved efficiently [3, 2, 18]. The optimization problem (2) is a *sparse coding* problem which finds the sparse codes α to represent x.

2.3 Dictionary Learning

Given sets of training data x_i, I want to estimate the dictionary D as well as the coefficients α_i for the sparse coding,

$$\{\hat{\alpha}_i, \hat{D}\} = \arg\min_{\alpha_i, D} \sum_{i}^{N} \lambda\|\alpha_i\|_1 + \|x_i - D\alpha_i\|_2^2, \tag{3}$$

where N is the number of training data, and λ is the parameter to control the sparsity of α_i. The problem is non-convex (bilinear in D and α_i). The standard approach [10] is alternating minimization, i.e., solving for α_i keeping D fixed and vice versa. The details of the algorithm can be found in [10, 7].

2.4 Coupled Dictionary Learning

Coupled dictionary learning is a method to learn a joint basis for compact representation of the data from two spaces [29, 30].

2.4.1 Standard Coupled Dictionary Learning (CDL)

Given sets of corresponding training pairs $\{x_i^{(1)}, x_i^{(2)}\}$, I can estimate the coupled dictionary D as well as the coefficients $\{\alpha_i\}$ by solving the minimization problem

$$\begin{aligned} \{\hat{D}, \{\hat{\alpha}_i\}\} &= \underset{D,\{\alpha_i\}}{\operatorname{argmin}} \sum_{i=1}^{N} \frac{1}{2} \left\| \begin{pmatrix} x_i^{(1)} \\ x_i^{(2)} \end{pmatrix} - \begin{pmatrix} D^{(1)} \\ D^{(2)} \end{pmatrix} \alpha_i \right\|_2^2 + \lambda\|\alpha_i\|_1 \\ &= \underset{D,\{\alpha_i\}}{\operatorname{argmin}} \sum_{i=1}^{N} \frac{1}{2}\|x_i - D\alpha_i\|_2^2 + \lambda\|\alpha_i\|_1, \end{aligned} \tag{4}$$

where N is the number of training data, and λ is the parameter to control the sparsity of α_i. We use a single α in (4) to enforce the correspondence of the dictionaries between two spaces.

2.4.2 Semi-coupled Dictionary Learning (SCDL)

In standard coupled dictionary learning, using only one set of coefficients imposes the strong assumption that coefficients of the representation of the two spaces are equal. However, this strong assumption sometimes does not hold. To relax this assumption, a semi-coupled dictionary learning is proposed [25],

$$\begin{aligned}\{\hat{D}^1, \hat{D}^2, \{\hat{\alpha}_i^1\}, \{\hat{\alpha}_i^2\}, \hat{W}\} = \underset{D^1, D^2, \{\alpha_i^1\}, \{\alpha_i^2\}, W}{\operatorname{argmin}} & \sum_{i=1}^{N} \frac{1}{2}\|x_i^1 - D^1\alpha_i^1\|_2^2 + \frac{1}{2}\|x_i^2 - D^2\alpha_i^2\|_2^2 \\ & + \lambda_1\|\alpha_i^1\|_1 + \lambda_2\|\alpha_i^2\|_1 + \gamma_1\|\alpha_i^2 - W\alpha_i^1\|_2^2 + \gamma_2\|W\|_F^2,\end{aligned} \tag{5}$$

where λ_1, λ_2, γ_1, γ_2 are regularization parameters. Distinct from CDL, W is a matrix to define a mapping between the coefficients in two spaces. Eq. (4) is a special case of Eq. (5) when W equals to the identity and α^1 equals α^2. Unlike for CDL the columns of the dictionaries D^1 and D^2 are normalized separately. Our experiments show that such a separate normalization is beneficial to jointly compute a basis for appearance differences and deformations. Eq. (5) 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.

3 CDL for Multi-modal Registration

Multi-modal registration is difficult due to the different appearances of the images in different modalities. By transforming the appearance of the image from one modality to another, the multi-modal registration problem can be simplified to a mono-modal one. The appearance transformation can be realized by image analogies. I propose an image analogies method based on coupled dictionary learning. I will first introduce how image analogies can be used in multi-modal registration.

3.1 Use of Image Analogies in Multi-modal Registration

For multi-modal image registration, I (i) reconstruct the “missing” analogous image and (ii) consistently denoise the given image to be registered with [11]. By denoising the target image using the learned dictionary for the target image from the joint dictionary learning step I obtain two consistently denoised images: the denoised target image and the predicted source image. The image registration is applied to the analogous image and the target image. The framework of the proposed method is illustrated in Fig. 1.

We have introduced coupled dictionary learning method in section 2.4. Here, I introduce the standard image analogies approach for appearance transformation.

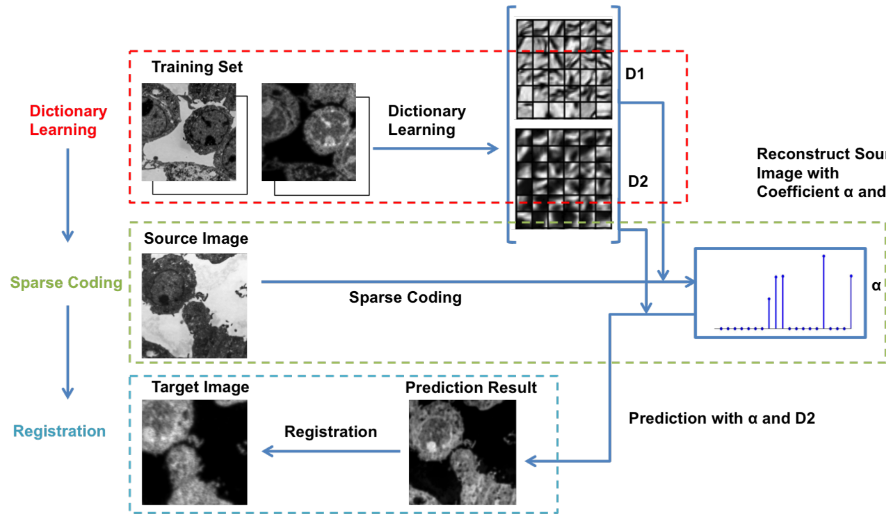
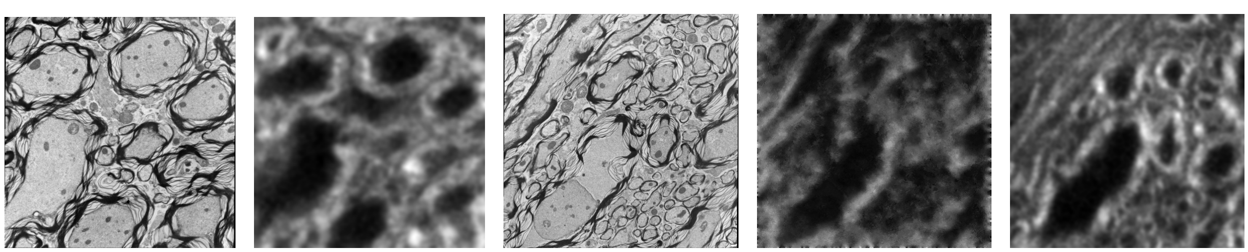


Figure 1: Flowchart of the proposed method. This method has three components: 1. dictionary learning: learning coupled dictionaries for both training images from different modalities; 2. sparse coding: computing sparse coefficients for the learned dictionaries to reconstruct the source image while at the same time using the same coefficients to transfer the source image to another modality; 3. registration: registering both transferred source image and target image.

3.2 Image Analogies for Appearance Transformation

The objective for image analogies is to create an image B' from an image B with a similar relation in appearance as a training image set (A, A') [13]. The standard image analogies algorithm achieves the mapping between B and B' by looking up best-matching patches for each image location between A and B which then imply the patch appearance for B' from the corresponding patch A' (A and A' are assumed to be aligned). These best patches are smoothly combined to generate the overall output image B'. Fig. 2 shows an example of image analogies.



(a) A:training TEM (b) A':training confocal (c) B:input TEM (d) B':output confocal (e) real output confocal

Figure 2: Result of Image Analogies: Based on a training set (A, A') an input image B can be transformed to B' which mimics A' in appearance. (e) is the real confocal image of (d).

Table 1: Prediction results for SEM/confocal images. Prediction is based on the proposed IA and standard IA methods, and I use sum of squared prediction residuals (SSR) to evaluate the prediction results. The p-value is computed using a paired t-test.

Method	mean	std	p-value
Proposed IA	$\mathbf{1.52 \times 10^5}$	5.79×10^4	**0.0002**
Standard IA	2.83×10^5	7.11×10^4	

3.3 CDL for Image Analogies

To avoid costly lookups and to obtain a more generalizable model with noise-reducing properties I propose an image analogies approach based on a sparse representation model and coupled dictionary learning.

The crucial part of the proposed method is to couple the dictionaries for the appearances of images from different modalities. By pre-registering training images, I obtain the correspondences between appearances of images in different modalities. Then corresponding image patch pairs $\{x_i^{(1)}, x_i^{(2)}\}$ are generated, and the coupled dictionary can be learned based on (4).

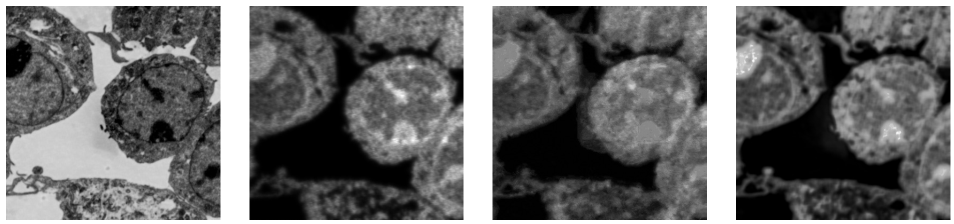
I formulate image analogies with a sparse representation model as,

$$\{\hat{u}^{(1)}, \hat{u}^{(2)}, \{\hat{\alpha}_i\}\} = \underset{\hat{u}^{(1)}, \hat{u}^{(2)}, \{\hat{\alpha}_i\}}{\operatorname{argmin}} \frac{\gamma}{2} \left\| u^{(1)} - f^{(1)} \right\|_2^2 + \frac{1}{2} \sum_{i=1}^{N} \left\| R_i \begin{pmatrix} u^{(1)} \\ u^{(2)} \end{pmatrix} - \begin{pmatrix} D^{(1)} \\ D^{(2)} \end{pmatrix} \alpha_i \right\|_2^2 + \lambda \|\alpha_i\|_1, \tag{6}$$

where I have corresponding multi-modal dictionaries $\{D^{(1)}, D^{(2)}\}$ and only one image $f^{(1)}$ is given and I am seeking a reconstruction of a denoised version of $f^{(1)}$, $u^{(1)}$, as well as the corresponding analogous denoised image $u^{(2)}$ (without the knowledge of $f^{(2)}$). $R_i(u)$ extracts the ith patch from image u. Note that there is only one set of coefficients α_i per patch, which indirectly relates the two reconstructions. The problem is convex (for given $D^{(i)}$) which allows to compute a globally optimal solution.

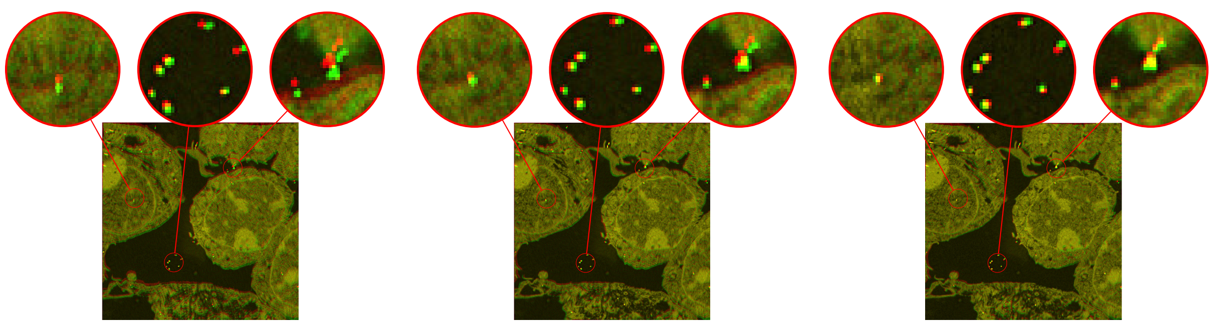
3.4 Results

The proposed method is tested on 2D correlative scanning electron microscopy (SEM)/confocal images with fiducials of a mouse brain. The test images are aligned with landmark based registration on the fiducials to obtain the gold standard results. Fig. 3 and Table 1 show the image analogies results. The analogous image based on the proposed method is more accurate compared with the standard IA method. Fig. 4 illustrates the registration results with and without image analogies based on the MI similarity measure. The proposed method provides more accurate registration results compared with the standard IA method and direct registration based on Fig. 4. More results can be found in [7].



(a) SEM Image (b) Confocal Image (c) Standard IA (d) Proposed IA

Figure 3: Results of estimating a confocal (b) from an SEM image (a) using the standard IA (c) and our proposed IA method (d).



(a) (b) (c)

Figure 4: Results of registration for SEM/confocal images using the MI similarity measure with direct registration (a), standard IA (b) and my proposed IA method (c) for b-spline registration. Some regions are zoomed in to highlight the distances between corresponding fiducials. The images show the compositions of the registered SEM images using the three registration methods (direct registration, standard IA and proposed IA methods) and the registered SEM image based on fiducials respectively.

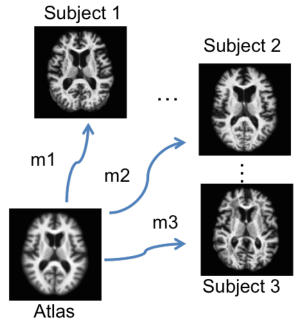


Figure 5: Illustration of training set and atlas. Here, m_i, $i = 1, \ldots, n$ are deformations generated by atlas construction [22].

4 SCDL for Deformation Estimation

In this section, I propose a semi-coupled dictionary learning method to estimate a deformation field based on image appearance. In particular, I explore how image appearance differences with respect to a common atlas image can be used to predict deformations represented by a joint basis. The joint basis allows us to investigate the distribution of image appearances and deformation jointly. Fig. 5 shows an example of training images for a set of subject images and their common atlas image.

The deformation prediction problem is challenging as it is difficult to model the relationship between image appearances and deformations. Also different deformation representation methods further complicate the modeling of the relationship. Moreover, little work has been done exploring the distribution of image appearance and deformations jointly. My proposed 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 (i) to establish if such bases can be learned, (ii) to establish if appearance differences can predict deformation differences, and (iii) to understand how appearance differences relate to deformations.

4.1 SCDL for Deformations and Appearances

In the proposed method, a coupled dictionary is learned using the SCDL method on image patches and corresponding deformation parameters. Although the dictionary learning method is similar to the method proposed in the Section 3, the goals of the dictionary learning are different. Here, the coupled dictionary can directly be used for deformation estimation based on image appearance. However, in the previous section, the coupled dictionary is applied to transform the image appearance from one modality to another. The deformation is estimated using standard image registration methods for a single modality.

4.2 Deformation Estimation

I first obtain D^1, D^2 and the linear mapping W from training data x_i^1 and x_i^2. Given a difference image $I = T - S$ (S is an input source image, and T is the common target/atlas image), similar to Eq. (5), I solve the following sparse coding problem,

$$\{\alpha_i^1\} = \underset{\alpha_i^1}{\operatorname{argmin}} \frac{1}{2} \|I_i^1 - D^1 \alpha_i^1\|_2^2 + \lambda_1 \|\alpha_i^1\|_1, \tag{7}$$

where I_i is a patch of I. Eq. (7) is sparse coding problem. 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 (for example, parameters for a B-spline transformation). After estimating ϕ_i from p_i for all the patches, I can determine the overall ϕ.

The method learns the relationship between appearance and deformation and obtains a compact representation of the data. The semi-coupled dictionary learning relaxes the strong coupling between the two spaces (here, appearance space and deformation space), thereby resulting in a better fit of the data distribution.

4.3 Results

The proposed method is tested on B-spline transformations [1]. I create a smoothed cross image as atlas. The synthetic subject images are generated by applying B-spline transformations. I generate coefficients for 3×3 control points by randomly sampling from a Gaussian distribution with $\sigma = 0.5$. Thus the parametrization for the deformation is the parameters of the control points for the B-spline transformation. The coupled dictionary is trained on difference images and B-spline transformation parameters. The dictionary is learned on the whole training image set. We have 1000 training images with image size 128×128, and learn the dictionaries with different numbers of items. We compared the SCDL method with the CDL method and the nearest neighbour (NN) method on deformation estimation for 100 test images. The results are shown in Table 2. The SCDL method performed better compared with the NN method and the CDL method. This demonstrates the power of our method to reduce the search space while at the same time maintaining good accuracy.

We also investigated other transformations from simple translation to deformable transformation parametrized by initial momentum [19]. More results on SCDL with different deformation parametrization methods can be found in [6]. Our experiments indicate that the learning procedure is able to capture a meaningful basis for the observed deformations.

Table 2: Statistics of registration results by predicting B-spline parameters for the B-spline transformation experiment. The numbers show the mean absolute errors (MAE) in pixels of predicted deformation to the ground truth. RAW indicates the images without any registration, NN indicates the method nearest neighbor search, GR represents the global regression method proposed in [9] and the CDL and SCDL represent coupled and semi-coupled dictionary learning methods respectively. The dictionary size is 500.

Method	median	mean	min	max	std
RAW	17.9776	19.2912	6.4132	45.1841	9.1304
NN	6.4063	7.1153	1.1399	12.2234	4.0948
GR	4.9261	5.2245	2.6279	10.2138	3.7925
CDL	3.6325	3.7531	0.724	8.5439	2.1158
SCDL	**3.0679**	**2.9254**	**0.3532**	**5.0559**	**1.4815**

5 CDL for GTPase Activities and Cell Protrusion

GTPases (Rac1, RhoA and Cdc42), a family of enzymes that can bind GTP, play important roles in cell dynamics [17, 20]. Cell dynamics are related to many biological processes such as tissue repair and regeneration, and disease progression of cancer [20]. Studying the relationship between GTPase activations and cell protrusion provides a way to understand the mechanism of cell dynamics thereby resulting in a better interpretation of biological processes.

I apply the coupled dictionary learning method to learn a joint basis for GTPase activations and cell movements in MEFs. Previous research has demonstrated correlations between GTPase activities and cell protrusion [17]. However, this previous approach focused on specific regions of MEFs, manually selected by experts. Our method captures the relationship between GTPase activation and cell protrusion by learning the coupled dictionary without hand-picking specific regions.

5.1 CDL for Joint GTPase Activities and Cell Protrusion Modeling

Before applying coupled dictionary learning, I need to prepare the data relating the information between GTPase activations and cell protrusion. Similar to the methods in [16, 17], the protrusion and retraction are defined as the trajectory of boundary points moving continuously perpendicular to the cell boundary[2]. I first track a set of cell boundary points. For each boundary point, I also record the mean intensity values of a small neighborhood, for example, a 5×5 neighborhood. The velocity for each boundary point can be obtained based on the orthogonal displacements with respect to the boundary and the imaging time interval of the biosensor. Learning coupled dictionary on data for single points usually lacks

[2]This definition of protrusion and retraction simplify the tracking process [16].

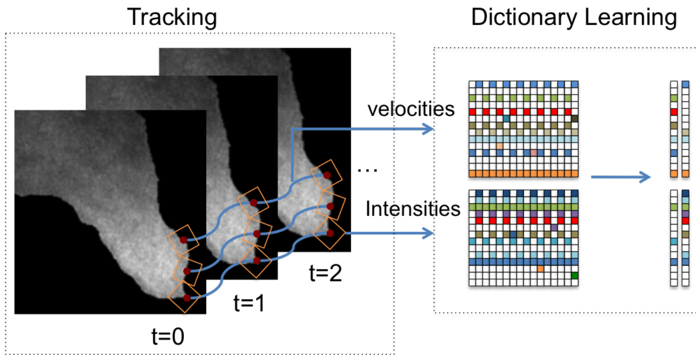


Figure 6: Flowchart for cell boundary tracking and coupled dictionary learning. The velocities of the boundary points (red points) are obtained by tracking points between consecutive frames and the mean intensities of the boundary points are acquired by averaging the intensities of the local neighborhood (orange squares). Coupled dictionary learning is applied to the obtained velocities and intensities for boundary points.

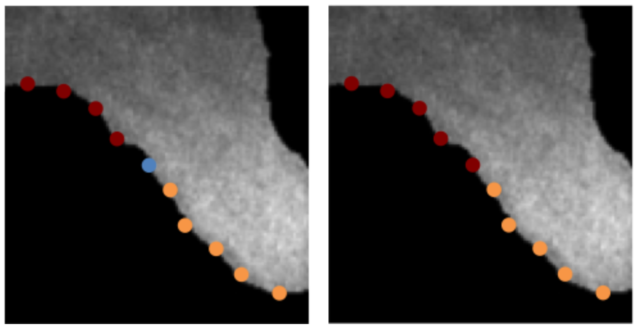
spatial consistency of neighboring points. Thus I also propose to learn the coupled dictionary on point "patches", i.e., over a small group of points (for example, 5 neighboring points together). Fig. 6 shows the flowchart of the proposed method.

5.2 Clustering of Cell Boundary Points based on GTPase Activation and Cell Protrusion Patterns

The coupled dictionary reveals the spatiotemporal relationship between GTPase activations and cell movements. Thus, the GTPase activation and corresponding protrusion information for each boundary point or point "patch" can be represented as a linear combination of the coupled dictionary atoms by solving a sparse coding problem in Section 2.2. The coefficients for the linear combination of the coupled dictionary atoms are new feature representations of the GTPase activation and cell protrusion patterns [28]. A large coefficient (in magnitude) indicates that the corresponding dictionary atom plays an important role in representing the data. The relationship of GTPase activation and cell protrusion for each boundary point can be studied based on the corresponding coefficients. To study the spatial interaction between neighboring boundary points or point "patches", one can partition points or "patches" into different sets so that similar points or "patches" are assigned to the same set. This can be achieved by clustering methods [14].

One of the most commonly used clustering methods is k-means clustering. K-means clustering divides the data into k different sets which aims to minimize the within-cluster errors [14]. However, this method requires specification of the number of clusters. It is usually difficult to find the optimal number of clusters, k. By hierarchical clustering [14], choosing a specific k is avoided by uncovering the hierarchical structure of the data [14]. Hierarchical clustering arranges the data in a tree structure which provides more information about the hierarchical relationship between different data points.

Sometimes single points may show different activation and protrusion patterns with their neighboring points, as a result, clustering on point “patches” can obtain more consistent results compared with clustering on single points. For example, in Fig. 7 (a), the blue point is assigned to a different cluster than its neighboring points. An example for a clustering result based on point “patches” is shown in Fig.7 (b).



(a) clustering on single points (b) clustering on point “patches”

Figure 7: Example of clustering results based on single points and point “patches”. Different colors indicate different clusters.

6 Robust Coupled Dictionary Learning (RCDL)

Coupled dictionary learning fails or provides inferior dictionary quality without sufficient correspondences between modalities in the training data. Fig. 8 shows an example of coupled dictionary learning for both perfect and imperfect corresponding image pairs. I propose a probabilistic model for coupled dictionary learning which distinguishes “noisy” training data (non-corresponding patch pairs) and refines the dictionary iteratively. The key component of the proposed method is defining the “noisy” training data with a confidence measure.

6.1 Confidence Measure

The confidence can be defined as a conditional probability $p(h|x_i)$. Given image patches $\{x_i\}_{i=1}^N$ I want to reconstruct them with our learned coupled dictionary. Here, h is the hypothesis of whether the reconstruction of x_i uses some ‘noisy’ dictionary atoms (i.e. non-corresponding dictionary atoms); $h = 1$ indicates that the reconstruction x_i uses ‘noisy’ dictionary atoms, while $h = 0$ means reconstructing x_i without ‘noisy’ dictionary atoms. Thus a high $p(h = 1|x_i)$ corresponds to high confidence that the reconstruction of x_i is not

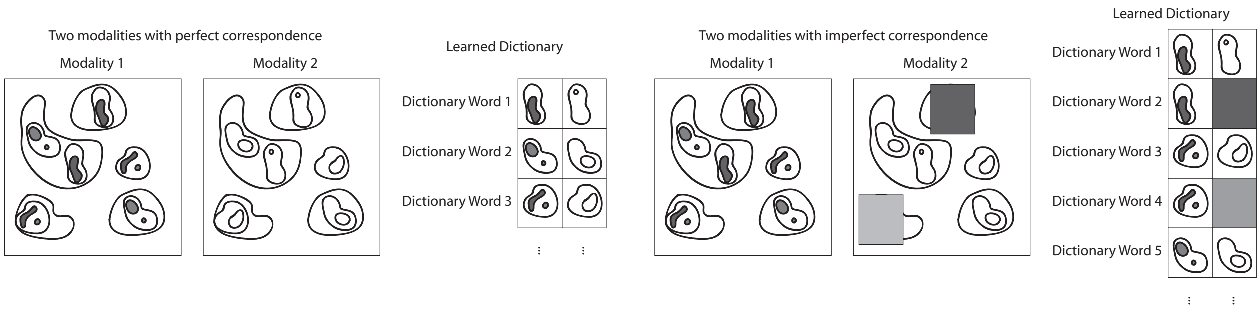


Figure 8: An illustration of perfect (left) and imperfect (right) correspondence between multi-modal images and their learned dictionaries. The imperfect correspondence (gray part in right images) could result in learning an imperfect dictionary (gray dictionary words) which is not desirable. Our goal is to *robustly* recover a compact dictionary of *corresponding* elements.

a good estimation because the 'noisy' dictionary atoms are used in the reconstruction and vice versa.

Applying Bayes' Rule, $P(h=1|x_i)$ can be represented as,

$$P(h=1|x_i) = \frac{P(x_i|h=1)P(h=1)}{P(x_i|h=1)P(h=1)+P(x_i|h=0)P(h=0)}. \tag{8}$$

Assuming the independence of each image patch x_i and that the pixels in each patch follow a Gaussian distribution, for $p(x_i|h)$ I assume

$$\begin{aligned} &p(x_i|h=1,\theta_1) = \mathcal{N}(x_i;\mu_1,\sigma_1^2), \\ &p(x_i|h=0,\theta_0;D,\alpha_i) = \mathcal{N}(x_i - D\alpha_i; 0, \sigma_0^2). \end{aligned} \tag{9}$$

The parameters I need to estimate are $\theta_1 = \{\mu_1, \sigma_1\}$ and $\theta_0 = \sigma_0$, as well as the prior probability $p(h)$, where $p(h=1) = \pi$ and $p(h=0) = 1-\pi$.

Based on the assumption of conditional independence of the random variable x_i given h and θ, maximum likelihood (ML) estimation are used for these parameters.

I propose an EM based algorithm which classifies the non-corresponding image patches and updates the coupled dictionary iteratively [5].

6.2 Results

The proposed method is tested on a synthetic example of SEM/confocal image pairs. Fig. 9 illustrates an example of SEM/confocal images. Non-corresponding patches with Gaussian noise are added to the images. The dictionary learned with RCDL and CDL are also shown in Fig 9 (in the right column and the middle column respectively). The dictionaries learned with RCDL exposes more information and less noise compared with the CDL method. More results on real data and multi-modal registration can be found in [5].

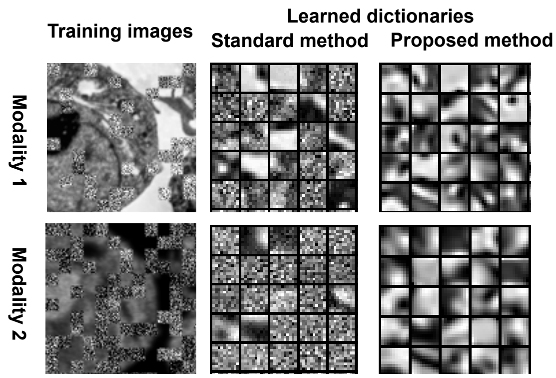


Figure 9: Dictionaries are learned from training SEM/confocal images (left column) with CDL and SCDL and shown in middle column and right column respectively).

7 Proposal Topics

I propose SCDL and RCDL in section 2.4.2 and section 6 respectively. Both of the methods aim to learn a coupled dictionary that can better explain the data compared with the standard CDL method in section 2.4. However, they are based on different assumptions. (i) For RCDL, the assumption is that the training data is deteriorated by "noise", for example, registration errors may introduce non-corresponding image patches in a training set. As a result, strong correspondence of the coupled dictionary is enforced by a single coefficient α in RCDL, however, the coupled dictionary is refined by removing non-corresponding training data. (ii) In SCDL, the assumption is that the relationship between the data from two spaces are not strongly coupled. Thus flexible coupling between the data from two different spaces is enabled in SCDL.

SCDL as a more general coupled dictionary learning method could be applied to multi-modal image registration. Thus I propose

- to apply SCDL to multi-modal image registration;
- to compare the results of both RCDL and SCDL in multi-modal applications;
- to analyze the relationship between GTPase activations and cell protrusions (incomplete in Section 5).

8 Summary

This proposal discussed issues related to modeling different coupled spaces with coupled dictionary learning. I first propose a coupled dictionary learning method for appearances of images from different modalities to transfer the appearance of one image from one modality to another. This simplifies the multi-modal registration problem to a mono-modal one. Moreover, I consider the problem of estimating the deformation field from a given subject image to a common atlas image by utilizing population information. A coupled dictionary is learned from the appearance differences between subject images and the atlas image and the corresponding deformations, and the deformation is estimated based on the relationship between the appearances and deformations from the coupled dictionary. Furthermore, I apply coupled dictionary learning to GTPase activation and cell edge movement for MEFs to reveal the common spatial and temporal patterns between GTPase activities and cell protrusion. I also propose a probabilistic model for robust coupled dictionary learning.

The proposed coupled dictionary learning methods (including RCDL and SCDL) essentially learn the relationship between the data in two spaces. RCDL and SCDL provide two different ways to better explain the data in two spaces compared with the standard CDL method. As a result, better prediction results (for both appearances and deformations) can be obtained based on these methods.

9 Research Status and Tentative Timeline

Table. 3 shows my research plan.

CDL for multi-modal registration	Section 3	Completed([7, 4])
CDL for deformation prediction	Section 4	Completed(accepted by ISBI 2015) In preparation to MEDIA
CDL for GTPase activation and cell protrusion modeling	Section 5	May 2015
Robust CDL	Section 6	Completed([5])
Other Proposal Topics	Section 7	June 2015
Proposal	______	January 2015
Oral Exam	______	February 2015
Defense	______	August 2015

Table 3: Reseach plan

References

[1] Fred L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on pattern analysis and machine intelligence*, 11(6):567–585, 1989.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Machine Learning*, 3(1):1–123, 2010.

[3] A.M. Bruckstein, D.L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM review*, 51(1):34–81, 2009.

[4] T. Cao, C. Zach, S. Modla, D. Powell, K. Czymmek, and M. Niethammer. Registration for correlative microscopy using image analogies. *Biomedical Image Registration*, pages 296–306, 2012.

[5] Tian Cao, Vladimir Jojic, Shannon Modla, Debbie Powell, Kirk Czymmek, and Marc Niethammer. Robust multimodal dictionary learning. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pages 259–266. Springer Berlin Heidelberg, 2013.

[6] Tian Cao, Nikhil Singh, Vladimir Jojic, and Marc Niethammer. Semi-coupled dictionary learning for deformation prediction. In *Submitted to ISBI*.

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

[8] J. Caplan, M. Niethammer, R.M. Taylor II, and K.J. Czymmek. The power of correlative microscopy: multi-modal, multi-scale, multi-dimensional. *Current Opinion in Structural Biology*, 2011.

[9] Chen-Rui Chou, Brandon Frederick, Gig Mageras, Sha Chang, and Stephen Pizer. 2d/3d image registration using regression learning. *Computer Vision and Image Understanding*, 117(9):1095–1106, 2013.

[10] M. Elad. *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer Verlag, 2010.

[11] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *Image Processing, IEEE Transactions on*, 15(12):3736–3745, 2006.

[12] D. Fronczek, C. Quammen, H. Wang, C. Kisker, R. Superfine, R. Taylor, DA Erie, and I. Tessmer. High accuracy fiona-afm hybrid imaging. *Ultramicroscopy*, 2011.

[13] A. Hertzmann, C.E. Jacobs, N. Oliver, B. Curless, and D.H. Salesin. Image analogies. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 327–340, 2001.

[14] Anil K Jain, M Narasimha Murty, and Patrick J Flynn. Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323, 1999.

[15] Minjeong Kim, Guorong Wu, Pew-Thian Yap, and Dinggang Shen. A general fast registration framework by learning deformation–appearance correlation. *Image Processing, IEEE Transactions on*, 21(4):1823–1833, 2012.

[16] M Machacek and G Danuser. Morphodynamic profiling of protrusion phenotypes. *Biophysical journal*, 90(4):1439–1452, 2006.

[17] Matthias Machacek, Louis Hodgson, Christopher Welch, Hunter Elliott, Olivier Pertz, Perihan Nalbant, Amy Abell, Gary L Johnson, Klaus M Hahn, and Gaudenz Danuser. Coordination of rho gtpase activities during cell protrusion. *Nature*, 461(7260):99–103, 2009.

[18] 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*, pages 689–696. ACM, 2009.

[19] Michael I Miller, Alain Trouvé, and Laurent Younes. Geodesic shooting for computational anatomy. *Journal of mathematical imaging and vision*, 24(2):209–228, 2006.

[20] Anne J Ridley, Martin A Schwartz, Keith Burridge, Richard A Firtel, Mark H Ginsberg, Gary Borisy, J Thomas Parsons, and Alan Rick Horwitz. Cell migration: integrating signals from front to back. *Science*, 302(5651):1704–1709, 2003.

[21] Yonghong Shi, Guorong Wu, Zhijian Song, and Dinggang Shen. Dense deformation reconstruction via sparse coding. In *Machine Learning in Medical Imaging*, pages 36–44. Springer, 2012.

[22] 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*, pages 1219–1222, 2013.

[23] Christian Wachinger and Nassir Navab. Manifold learning for multi-modal image registration. *11st British Machine Vision Conference (BMVC)*, 2010.

[24] Qian Wang, Minjeong Kim, Guorong Wu, and Dinggang Shen. Joint learning of appearance and transformation for predicting brain mr image registration. In *Information Processing in Medical Imaging*, pages 499–510. Springer, 2013.

[25] 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*, pages 2216–2223, 2012.

[26] Wolfgang Wein, Shelby Brunke, Ali Khamene, Matthew R Callstrom, and Nassir Navab. Automatic ct-ultrasound registration for diagnostic imaging and image-guided intervention. *Medical image analysis*, 12(5):577, 2008.

[27] W.M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical image analysis*, 1(1):35–51, 1996.

[28] 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*, 98(6):1031–1044, 2010.

[29] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang. Coupled dictionary training for image super-resolution. *Image Processing, IEEE Transactions on*, 21(8):3467–3478, 2012.

[30] Jianchao Yang, Zhaowen Wang, Zhe Lin, Xianbiao Shu, and Thomas Huang. Bilevel sparse coding for coupled feature spaces. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2360–2367. IEEE, 2012.

[31] S. Yang, D. Kohler, K. Teller, T. Cremer, P. Le Baccon, E. Heard, R. Eils, and K. Rohr. Nonrigid registration of 3-d multichannel microscopy images of cell nuclei. *Image Processing, IEEE Transactions on*, 17(4):493–499, 2008.