COUPLED DICTIONARY LEARNING FOR IMAGE ANALYSIS

Tian Cao

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill
2016

Approved by:

Marc Niethammer

Alexander C. Berg

Klaus M. Hahn

Martin Styner

Vladimir Jojic

ABSTRACT

Tian Cao: COUPLED DICTIONARY LEARNING FOR IMAGE ANALYSIS
(Under the direction of Marc Niethammer)

Modern imaging technologies provide 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 instance, relating protein locations from fluorescence microscopy and the protein structures from electron microscopy. These problems are challenging due to the difficulties in modeling the relationship between the information from different modalities.

In this dissertation, a coupled dictionary learning based image analogy method is first introduced to synthesize images in one modality from images in another. As a result, using my method multi-modal registration (for example, registration between correlative microscopy images) is simplified to a mono-modal one. Furthermore, a semi-coupled dictionary learning based framework is proposed to estimate deformations from image appearances. Moreover, a coupled dictionary learning method is explored to capture the relationship between GTPase activations and cell protrusions and retractions. Finally, a probabilistic model is proposed 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.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Marc Niethammer, for introducing me to the world of medical image analysis and for keeping me motivated with his great enthusiasm for the subject. I have continuously benefited from his knowledge, his encouragement and his personal guidance.

I would also like to thank my committee members, Alexander C. Berg, Klaus M. Hahn, Martin Styner and Vladimir Jojic, for their feedback and advice.

Additionally, I would like to thank my labmates for their help and support: Liang Shan, Yang Huang, Istvan Csapo, Yi Hong, Xiao Yang, Heather Couture, Xu Han, Nikhil Singh, Thomas Polzin.

Last but not least, I would like to thank my parents and my girlfriend for all of their encouragement and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DL	Dictionary Learning
CDL	Coupled Dictionary Learning
SCDL	Semi-coupled Dictionary Learning
RCDL	Robust Coupled Dictionary Learning
IA	Image Analogies
NN	Nearest Neighbor
GR	Global Regression
SSD	Sum of Squared Differences
MI	Mutual Information
EM	Expectation Maximization
TEM	Transmission Electron Microscopy
MEF	Mouse Embryonic Fibroblast
MR	Magnetic Resonance
PCA	Principal Component Analysis
GTP	Guanosine Triphosphate

CHAPTER 1: INTRODUCTION

1.1 Motivations

With the development of new imaging technologies, we can visualize various objects to explore information ranging from molecular structures of cells to tissue of the human body. Different imaging modalities provide distinct information about the objects. For example, in the context of correlative microscopy, which combines different microscopy technologies such as conventional light-, confocal- and electron transmission microscopy (Caplan et al., 2011), protein locations can be revealed through fluorescence microscopy, while protein structures can be observed through electron microscopy (Caplan et al., 2011).

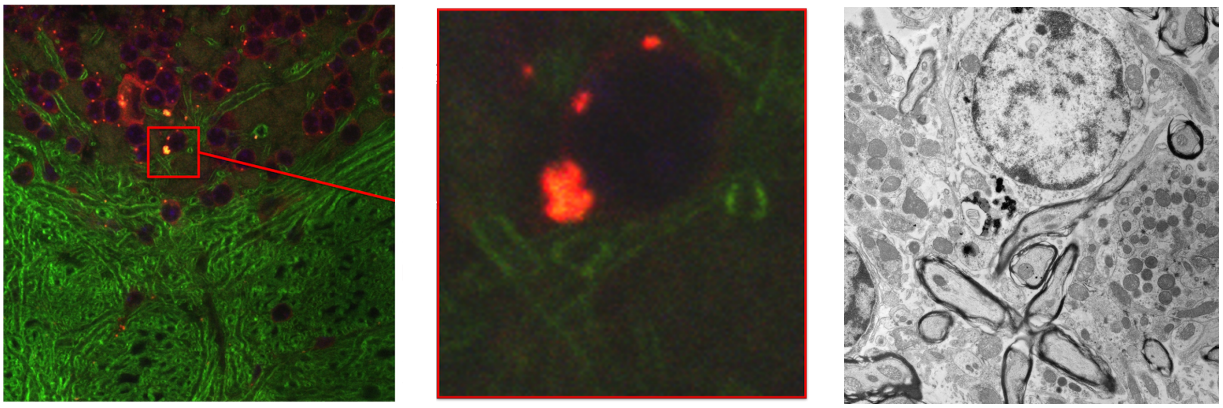
Many image analysis applications require relating the information from different modalities or sources. For example, joint analysis of correlative microscopic images needs registration of images from different modalities as illustrated in Figure 1.1; estimating the deformation of an image with respect to an atlas requires modeling the relationship between deformations and image appearances as illustrated in Figure 1.2; investigating the coordination of GTPase activities and cell protrusions of mouse embryonic fibroblasts (MEFs) requires establishing the spatio-temporal correspondences between GTPase activations and cell movements as illustrated in Figure 1.3.

The similarity between these applications is that they require us to *model the relationship of data from different spaces*.

1.1.1 Multi-modal image registration for correlative microscopy

Correlative microscopy is a methodology combining the functionality of light microscopy with the high resolution of electron microscopy and other microscopy technologies. One of the most important steps to combining the information from different types of microscopy is to perform registration between two or more microscopic images (Cao et al., 2013b). Image registration is the

process of estimating spatial transformations between images (to align them). Registration requires a transformation model and a similarity measure. Direct intensity based similarity measures (such as the sum of squared intensity differences (SSD)) is typically not applicable to multi-modal image registration as image appearances/intensities differ. While a similarity measure such as mutual information (MI) can account for image appearance differences (as MI does not rely on direct image intensity comparisons), it is a very generic measure of image similarity not adapted to a particular application scenario. In general, as images in correlative microscopy differ severely, measuring image similarity is difficult and therefore image registration is challenging. Figure 1.1 shows an example of correlative microscopy.



(a) Confocal Microscopic Image (b) Resampling of Region in (a) (c) TEM Image

Figure 1.1: Example of Correlative Microscopy. (a) is a stained confocal brain slice, where the red box shows a selected region and (b) is a resampled image of the boxed region in (a). The goal is to align (b) to (c).

1.1.2 Deformation estimation from appearances

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, to accelerate the registration process or to initialize other registration methods (Chou et al., 2013; Wang et al., 2013). Figure 1.2 shows an example of an atlas image and a set of corresponding subject images.

The deformation estimation problem is challenging as it is difficult to model the relationship between image appearances and deformations. For example, the relation between appearances and deformations could be highly nonlinear.

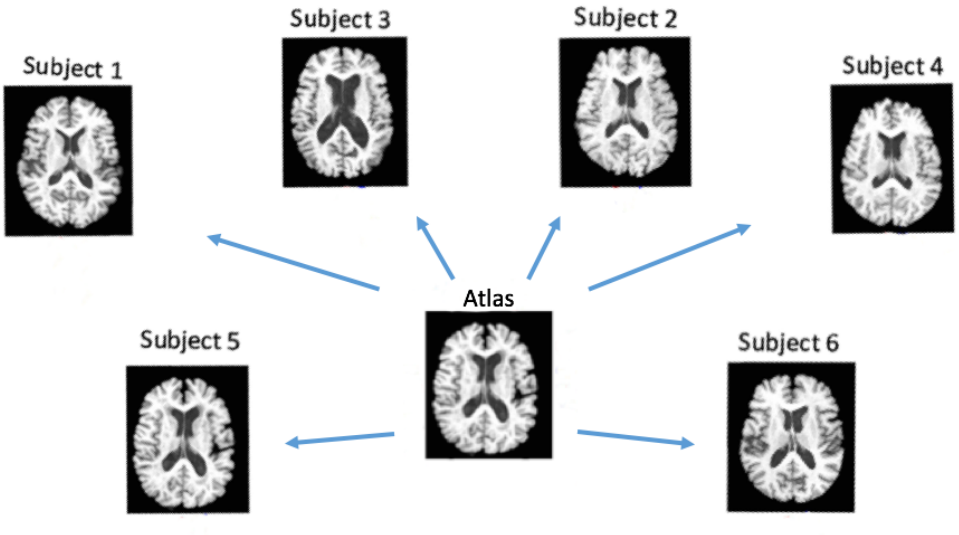


Figure 1.2: Example of atlas image and the corresponding subject images.

1.1.3 Coordination between GTPase activities and cell movements

GTPases play key roles in controlling cell dynamics (Ridley et al., 2003; Machacek et al., 2009). Exploring the spatiotemporal coordination of GTPase activations and cell protrusions is crucial for the understanding of cell dynamics (Machacek et al., 2009; Ridley et al., 2003). GTPases (Rac1, RhoA and Cdc42) are a family of enzymes that can bind GTP. GTPase activities can be studied with biosensor imaging (Machacek et al., 2009). However, exploring the relationship between GTPase activations and cell protrusions 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. Figure 1.3 shows an example of GTPase activations of a cell.

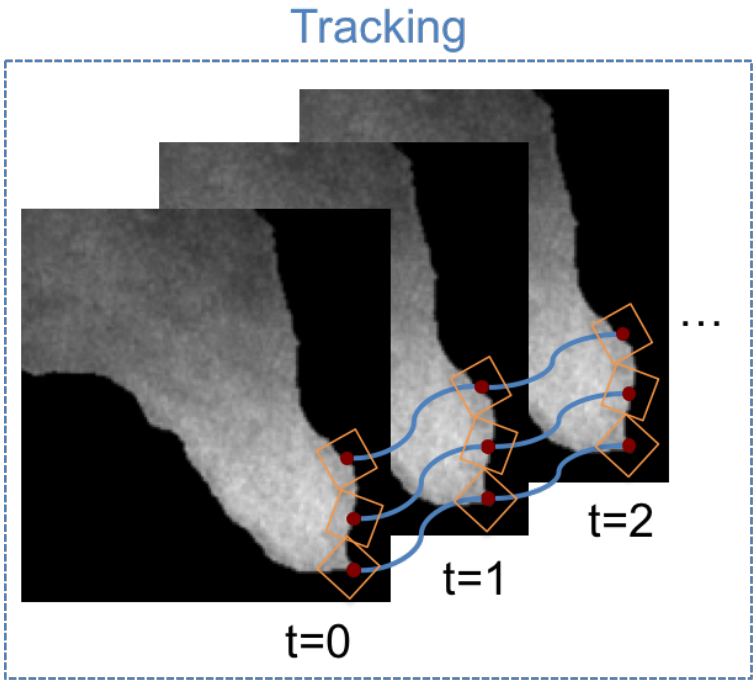


Figure 1.3: Example of GTPase activations of a cell. Intensities represent the GTPase activations. Some boundary points are tracked across different time points to extract the corresponding activations and cell movements (velocities). Refer to Chapter 5 for more details.

1.1.4 Coupled dictionary learning

In this dissertation, coupled dictionary learning based approaches are proposed to *learn* these relationships[1] from given data. Dictionary learning is a method to learn a basis to represent data (Mairal et al., 2009; Kreutz-Delgado et al., 2003). Typically over-complete dictionaries are used where the number of basis vectors is larger than the dimensionality of the data. Similarly, coupled dictionary learning learns a coupled basis for the data from two coupled spaces. Coupled spaces can for example be the two appearance spaces from two different image modalities. A coupled dictionary can then relate appearances between the two spaces.

1.1.5 Robust coupled dictionary learning

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

[1]For example, the relationships between appearances of images from different modalities, between appearances and deformations and between GTPase activations and cell protrusions.

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, by pre-registering training images before dictionary learning. Errors can be introduced during the process of establishing these correspondence.

Thus, a robust dictionary learning method under a probabilistic model is finally introduced in Chapter 6. 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.

1.2 Contributions

The main contributions of my work include:

1. a sparse representation and a coupled dictionary learning-based image analogy method to convert and thereby simplify a multi-modal registration problem to a mono-modal one;
2. a general framework for deformation estimation using appearance information based on coupled dictionary learning;
3. a framework for relating the spatio-temporal patterns between GTPase activations and cell protrusion of mouse embryonic fibroblasts (MEFs) based on coupled dictionary learning;
4. a robust coupled dictionary learning method based on a probabilistic model which discriminates between corresponding and non-corresponding patches automatically.

1.3 Thesis statement

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 1) appearance differences of different imaging modalities, 2) dependencies between image appearance and deformation, and 3) spatio-temporal 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.

1.4 Outline

The remainder of the document is organized as follows:

Chapter 2 introduces some background on related methods of this dissertation; Chapter 3 describes the image analogy based multi-modal registration; Chapter 4 explains the coupled dictionary learning method for both appearances and deformations; Chapter 5 presents an application of coupled dictionary learning for modeling GTPase activation and cell protrusion for MEFs; Chapter 6 proposes a robust multi-modal dictionary learning method; Chapter 7 concludes with a discussion of contributions and some possible future work.

CHAPTER 2: BACKGROUND

This chapter provides necessary background and approaches related to this dissertation. Section 2.1 introduces image registration. Section 2.2 introduces the standard image analogy method. Section 2.3 provides an introduction to the sparse representation model and its applications in image processing and computer vision. Section 2.4 discusses the dictionary learning method and the numerical solution of the associated learning problem. In Section 2.5, coupled dictionary learning is introduced to capture the relationship between data in two spaces.

2.1 Image registration

Image registration estimates a spatial 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, $y = \phi(x)$. Let S and T represent the source (moving) image and the target (static) image respectively. The goal of image registration is to estimate the transformation $\phi(\cdot)$ for the entire source image that transforms it to match the subject's image in the sense that, the appearance at a pixel y in the deformed source image, $\tilde{S}(y) = S(\phi^{-1}(y))$ best matches with the target image, $T(y)$. Figure 2.1 shows the framework for standard image registration (Ibanez et al., 2003; Hill et al., 2001; Maintz and Viergever, 1998). In this framework, image registration is treated as an optimization problem. The interpolator is used to estimate the intensity values of the source image at non-grid locations (Ibanez et al., 2003), while the distance metric d measures the difference between the transformed source image and the target image. The metric d provides a quantitative measure which can be optimized by the optimizer in the search space defined by the transformation parameters[1] (Ibanez et al., 2003).

[1] more details can be found in Section 4.3.

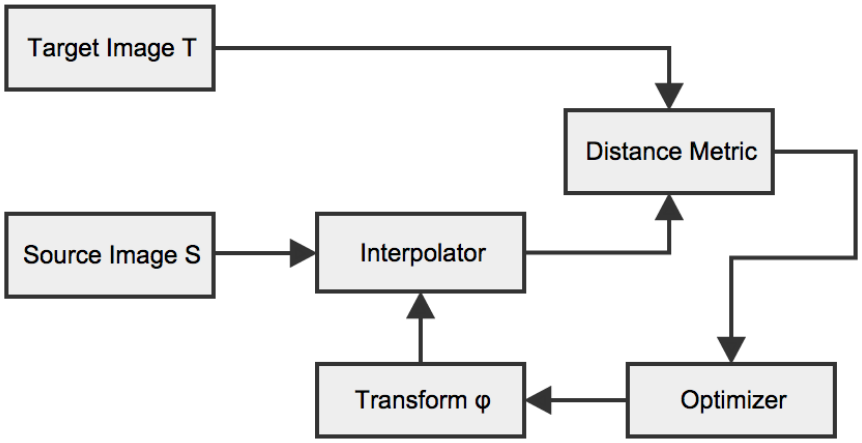


Figure 2.1: Framework of the standard image registration.

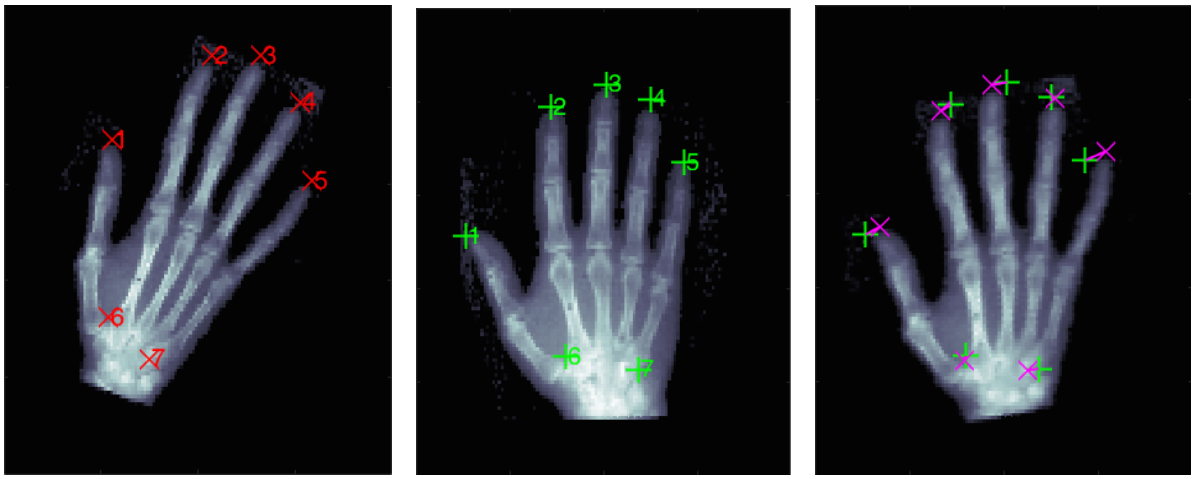
There are many image registration methods (Maintz and Viergever, 1998; Zitova and Flusser, 2003). One simple method is landmark-based registration. Here landmarks are a set of points in both source and target images. The idea of landmark-based registration is to estimate a transformation that minimizes the distance between corresponding landmarks in source and target images. Let $s_i = [s_{xi}, s_{yi}]$ denote the position of the ith landmark in the source image and $t_i = [t_{xi}, t_{yi}]$ the position of the corresponding landmark in the target image, where $i = 1, \ldots, n$, n is the number of landmarks. Landmark-based registration solves the following minimization problem,

$$\min_{\phi} d(s_i, t_i, \phi) = \min_{\phi} \sum_{i=1}^{n} \|\phi(s_i) - t_i\|_2^2, \tag{2.1}$$

where d is the distance between corresponding landmarks, here I use the sum of squared Euclidean distances as an example.

If no landmarks are available, intensity-based registration methods can be applied to the source and target images. Intensity-based methods solve,

$$\min_{\phi} d(S, T, \phi) = \min_{\phi} \sum_{x \in \Omega} \|S(\phi^{-1}(x)) - T(x)\|_2^2, \tag{2.2}$$



(a) source image (b) target image (c) transformed source image

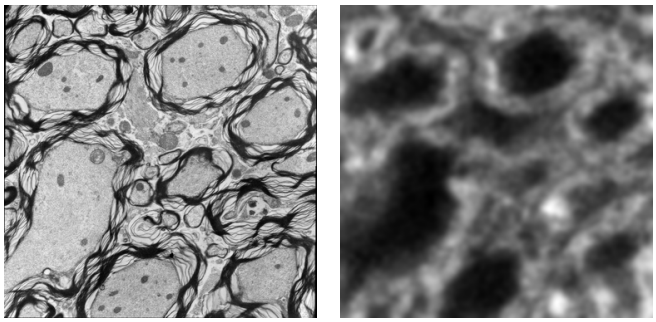
Figure 2.2: Example of landmark-based registration. The landmarks are superimposed on (a) source image and (b) target image. The transformed source image and corresponding landmarks are shown in (c). The images are from (Modersitzki, 2009).

where Ω is the domain of the image, d is the distance which measures the similarity between source and target images. Here I use the sum of squared difference (SSD) as an example. Another commonly used distance measure is mutual information (MI) (Wells et al., 1996).

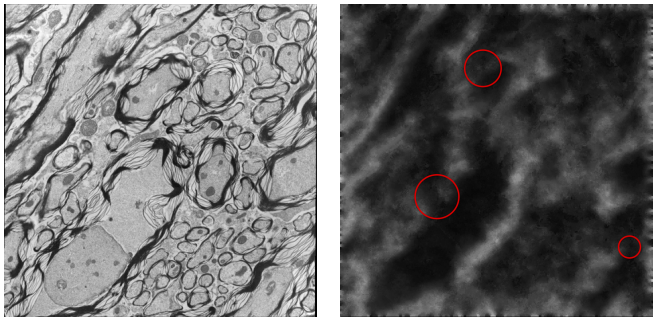
Which similarity measure to pick for image registration depends on the image pairs to be registered. If they are from the same modality, SSD may be a good choice. For multi-modal approaches more flexible similarity measures such as MI are typically used. However, for image pairs with drastically different appearances such measures may also not be ideal. In this thesis I therefore investigate creating a custom image similarity term via the process of image synthesis which I will discuss next.

2.2 Image analogy

Image analogies, first introduced in (Hertzmann et al., 2001), have been widely used in texture synthesis. Image analogy allows to relate the appearance between two images. In this method, a pair of images A and A' is provided as training data, where A' is a "filtered" version of A. The "filter" is learned from A and A' and is later applied to a different image B in order to generate an "analogous" filtered image B'. Fig. 2.3 shows an image analogy example.



(a) A:training TEM (b) A':training confocal



(c) B:input TEM (d) B':output confocal

Figure 2.3: 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. The red circles in (d) show inconsistent regions.

The standard image analogy algorithm achieves the mapping between B and B' by looking up best-matching image 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). Examples for image patches are shown in Figure 2.4. These best-matched patches are smoothly combined to generate the overall output image B'. Figure 2.4 also shows the feature matching between the points in A, A' and B, B'. The algorithm description is presented in Algorithm 1.

2.2.1 Nearest neighbor search

For multi-modal image registration in Chapter 3, the image analogy approach can be used to transfer a given image from one modality to another using the trained "filter". Then the multi-modal image registration problem simplifies to a mono-modal one. However, since this method uses a nearest neighbor (NN) search of the image patch centered at each pixel, the resulting images are usually noisy because the ℓ_2 norm based NN search does not preserve the local consistency well (see Figure 2.3 (d)) (Hertzmann et al., 2001). Another problem is the size of the search space, i.e.

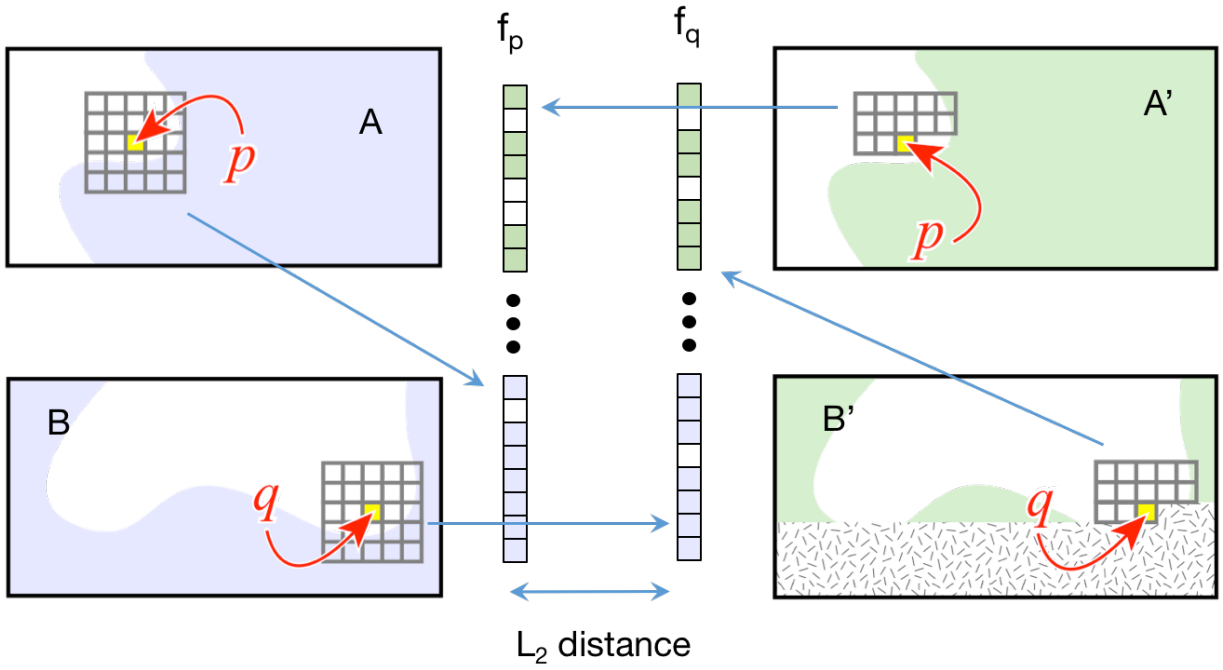


Figure 2.4: Feature matching in the image analogy method: for a point p in training set (A, A'), the feature f_p for p is extracted from the image patches centered at p; similarly for a point q in the input image B and the corresponding sythesized B', the feature f_q is extracted from the image patches centered at q. The feature vector concatenates the intensity values for the image patches centered at a point p in both A and A' or B and B'. Suppose p is the closest point of q based on the ℓ_2 distance of the feature vectors, the synthesized intensity at point q in B' is set to the intensity value at p in A'.

the number of patches in the NN search, is linearly related to the size of the training images. As a result, NN search is time consuming for large training sets. I introduce a sparse representation model to address these problems in Section 2.3.

Algorithm 1 Image Analogies.

Input:
Training images: A and A';
Source image: B.
Output:
'Filtered' source B'.
Construct Gaussian pyramids for A, A' and B;
Generate features for A, A' and B;
for each level l starting from coarsest **do**
  for each pixel $q \in B'_l$, in scan-line order **do**
    Find best matching pixel p of q in A_l and A'_l;
    Assign the value of pixel p in A' to the value of pixel q in B'_l;
    Record the position of p.
  end for
end for
Return B'_L where L is the finest level.

2.3 Sparse representation

Sparse representation is a powerful model for representing and compressing signals (Wright et al., 2010; Huang et al., 2011b). It represents a signal as a combination (usually linear) of a few fixed basis signals from a typically over-complete dictionary (Bruckstein et al., 2009). It has been successfully applied in many computer vision applications such as for object recognition and classification in (Wright et al., 2009; Huang and Aviyente, 2007; Huang et al., 2011a; Zhang et al., 2012a,b; Fang et al., 2013; Cao et al., 2013a). 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 (Elad, 2010):

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \ \|\alpha\|_0, \quad \text{s.t.} \ \|x - D\alpha\|_2 \leq \epsilon, \tag{2.3}$$

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 Equation (2.3) is an NP-hard problem. One possible solution of this problem is based on a relaxation that replaces $\|\cdot\|_0$ by $\|\cdot\|_1$, where $\|\cdot\|_1$ is the 1-norm of a vector, resulting in the optimization problem,

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \ \|\alpha\|_1, \quad \text{s.t.} \ \|x - D\alpha\|_2 \leq \epsilon. \tag{2.4}$$

For a suitable choice of λ, an equivalent problem is given by Equation (2.4)

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \ \|x - D\alpha\|_2^2 + \lambda\|\alpha\|_1, \tag{2.5}$$

which is a convex optimization problem that can be solved efficiently (Bruckstein et al., 2009; Boyd et al., 2010; Mairal et al., 2009). The optimization problem Equation (2.5) is a *sparse coding* problem which finds the sparse codes α to represent x. Based on the sparse representation model, an over-complete basis, i.e. a dictionary, can be learned for the reconstruction of data in some space.

2.4 Dictionary learning

Dictionary learning plays a key role in many applications using sparse models. As a result, many dictionary learning methods have been introduced in recent literature (Aharon et al., 2006; Yang et al., 2010; Mairal et al., 2008; Monaci et al., 2007). In (Aharon et al., 2006), a dictionary is learned for image denoising, while in (Mairal et al., 2008), supervised dictionary learning is performed for classification and recognition tasks. Given sets of training data x_i, the goal is to estimate the dictionary D as well as the coefficients α_i for the sparse coding problem,

$$\{\hat{\alpha}_i, \hat{D}\} = \underset{\alpha_i, D}{\operatorname{argmin}} \ \sum_{i=1}^{N} \|x_i - D\alpha_i\|_2^2 + \lambda\|\alpha_i\|_1, \tag{2.6}$$

where N is the number of training datasets, and λ is the parameter to control the sparsity of α_i. The ℓ_1 regularity term induces sparsity in 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 $D = \{d_1, d_2, ..., d_q\} \in \mathbb{R}^{p \times q}$.

2.4.1 Data preprocessing

Let X be the data matrix $X = \{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^{p \times n}$, each column in X represents a data vector, while each row of X represents a feature. Usually X requires preprocessing before dictionary learning. There are three commonly used data preprocessing methods.

Mean subtraction : Here, for each feature in the data matrix its mean across all data points is subtracted. The geometric interpretation of mean subtraction is centering the data around the origin for every feature dimension.

Feature Scaling : Here, each feature in the data matrix is divided by its standard deviation across all data points *after* mean subtraction. It is usually applied to data where each feature has a different scale but equal importance to the learning algorithm. For image patches, it is not necessary to apply feature scaling because the relative scales of pixels are approximately equal (usually between 0 and 255).

Normalization : Here, each data vector is normalized to unit norm by division with its ℓ_2 norm.

Note that it is not necessary to apply all the preprocessing steps to the data. Choosing appropriate data preprocessing methods is based on the applications. I will give an example in Section 3.3.3.

2.4.2 Numerical solution

The optimization problem in Equation (2.6) is non-convex (bilinear in D and α_i). The standard approach (Elad, 2010) is alternating minimization, i.e., solving for α_i keeping D fixed and vice versa. By fixing D, we first solve the sparse coding problems,

$$\hat{\alpha}_i = \underset{\alpha_i}{\operatorname{argmin}} \sum_{i}^{N} \|x_i - D\alpha_i\|_2^2 + \lambda \|\alpha_i\|_1. \tag{2.7}$$

Following (Li and Osher, 2009) I use a coordinate descent algorithm to solve Equation (2.7). The coordinate descent algorithm to solve Equation (2.7) is described in Algorithm 2. This

algorithm minimizes Equation (2.7) with respect to one component of α in one step, keeping all other components constant. This step is repeated until convergence.

Algorithm 2 Coordinate Descent for Sparse Coding

Input: $\alpha = 0,\ \lambda > 0,\ \beta = D^T x$
Output: α
while not converged **do**
1. $\tilde{\alpha} = S_\lambda(\beta)$[1];
2. $j = \text{argmax}_i |\alpha_i - \tilde{\alpha}_i|$, where i is the index of the component in α and $\tilde{\alpha}$;
3. $\alpha_i^{k+1} = \alpha_i^k,\ i \neq j$, and $\alpha_j^{k+1} = \tilde{\alpha}_j$;
4. $\beta^{k+1} = \beta^k - |\alpha_j^k - \tilde{\alpha}_j|(D^T D)_j$, and $\beta_j^{k+1} = \beta_j^k$, where $(D^T D)_j$ is the jth column of $(D^T D)$.

end while

After solving Equation (2.7), I can fix α_i and then update D. Now the optimization of Equation (2.6) can be changed to

$$\hat{D} = \underset{D}{\text{argmin}} \sum_{i=1}^{N} \|x_i - D\alpha_i\|_2^2. \tag{2.8}$$

The closed-form solution of Equation (2.8) is as follows,

$$D = (\sum_{i=1}^{N} x_i \alpha_i^T)(\sum_{i=1}^{N} \alpha_i \alpha_i^T)^{-1}. \tag{2.9}$$

The columns are normalized according to

$$d_j = d_j / \max(\|d_j\|_2, 1),\ j = 1, ..., m, \tag{2.10}$$

where $D = \{d_1, d_2, ...d_m\} \in \mathbb{R}^{nxm}$. I iterate the optimization with respect to D and α_i to convergence.

2.5 Coupled dictionary learning

Coupled dictionary learning is a method to learn a joint basis for compact representation of the data from two spaces (Yang et al., 2012a,b). In (Monaci et al., 2007), a coupled dictionary

[2] $S_a(v)$ is soft thresholding operator where $S_a(v) = (v - a)_+ - (-v - a)_+$.

is learned from audio-visual data. A coupled dictionary which has two parts corresponding to different modalities, can be also applied to super-resolution (Yang et al., 2010) and multi-modal image registration (Cao et al., 2012).

2.5.1 Standard coupled dictionary learning (CDL)

Given sets of corresponding training pairs $\{x_i^{(1)}, x_i^{(2)}\}$, the coupled dictionary learning problem can be formulated as

$$\{\hat{\tilde{D}}, \hat{\alpha}\} = \underset{\tilde{D},\alpha}{\operatorname{argmin}} \sum_{i=1}^{N} \frac{1}{2}\|\tilde{x}_i - \tilde{D}\alpha_i\|_2^2 + \lambda\|\alpha_i\|_1, \tag{2.11}$$

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. Note that there is only one set of coefficients α_i per datum, which enforces correspondence of the dictionaries between the two spaces. The numerical solution for standard CDL is the same as the solution mentioned in Section 2.4.2.

2.5.2 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 in (Wang et al., 2012),

$$\begin{aligned}\{\hat{D}^{(1)}, \hat{D}^{(2)}, \{\hat{\alpha}_i^{(1)}\}, \{\hat{\alpha}_i^{(2)}\}, \hat{W}\} = \underset{D^{(1)},D^{(2)},\{\alpha_i^{(1)}\},\{\alpha_i^{(2)}\},W}{\operatorname{argmin}} \sum_{i=1}^{N} \frac{1}{2}\|x_i^{(1)} - D^{(1)}\alpha_i^{(1)}\|_2^2 + \\ \frac{1}{2}\|x_i^{(2)} - D^{(2)}\alpha_i^{(2)}\|_2^2 + \lambda_1\|\alpha_i^{(1)}\|_1 + \lambda_2\|\alpha_i^{(2)}\|_1 + \gamma_1\|\alpha_i^{(2)} - W\alpha_i^{(1)}\|_2^2 + \gamma_2\|W\|_F^2,\end{aligned} \tag{2.12}$$

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. Equation (2.11) is a special case of Equation (2.12) when W equals the identity and $\alpha^{(1)}$ equals $\alpha^{(2)}$. Unlike for CDL the columns of the dictionaries $D^{(1)}$ and $D^{(2)}$ are normalized separately. In Chapter 4, our experiments show that such a separate

normalization is beneficial to jointly compute a basis for appearance differences and deformations. Equation (2.12) is not convex with respect to $D^{(1)}$, $D^{(2)}$, $\alpha^{(1)}$, $\alpha^{(2)}$, W jointly, however, it is convex with respect to each of them when others are fixed.

2.5.3 Numerical solution

Equation (2.12) can be solved by an iterative algorithm updating $D^{(1)}$, $D^{(2)}$, W alternatingly. I initialize W as identity matrix and $D^{(1)} = \{d_1^{(1)}, \ldots, d_m^{(1)}\}$ and $D^{(2)} = \{d_1^{(2)}, \ldots, d_m^{(2)}\}$ as m random $x_i^{(1)}$ and $x_i^{(2)}$ pairs. With W, $D^{(1)}$, $D^{(2)}$ fixed, I solve the following sparse coding problems with respect to $\alpha_i^{(1)}$ and $\alpha_i^{(2)}$,

$$\begin{aligned}
&\min_{\alpha_i^{(1)}} \sum_{i=1}^{N} \frac{1}{2}\|x_i^{(1)} - D^{(1)}\alpha_i^{(1)}\|_2^2 + \gamma_1\|\alpha_i^{(2)} - W\hat{\alpha}_i^{(1)}\|_2^2 + \lambda_1\|\alpha_i^{(1)}\|_1, \\
&\min_{\alpha_i^{(2)}} \sum_{i=1}^{N} \frac{1}{2}\|x_i^{(2)} - D^{(2)}\alpha_i^{(2)}\|_2^2 + \gamma_1\|\hat{\alpha}_i^{(2)} - W\alpha_i^{(1)}\|_2^2 + \lambda_2\|\alpha_i^{(2)}\|_1,
\end{aligned} \tag{2.13}$$

where $\hat{\alpha}_i^{(1)}$ and $\hat{\alpha}_i^{(2)}$ are results obtained from the previous iteration. Equation (2.13) are lasso problems which can be solved by Algorithm 2 or other existing ℓ_1 solvers such as coordinate descent (Friedman et al., 2007), FISTA (Beck and Teboulle, 2009) and LARS (Efron et al., 2004).

By fixing W, $\alpha_i^{(1)}$ and $\alpha_i^{(2)}$, based on Equation (2.11), I update $D^{(1)}$ and $D^{(2)}$ by solving

$$\begin{aligned}
&\min_{D^{(1)}, D^{(2)}} \sum_{i=1}^{N} \frac{1}{2}\|x_i^{(1)} - D^{(1)}\alpha_i^{(1)}\|_2^2 + \frac{1}{2}\|x_i^{(2)} - D^{(2)}\alpha_i^{(2)}\|_2^2, \\
&\text{s.t. } \|d_j^{(1)}\|_2 \leq 1,\ \|d_j^{(2)}\| \leq 1,\ j = 1, \ldots, m,
\end{aligned} \tag{2.14}$$

where $D^{(1)} = \{d_1^{(1)}, \ldots, d_m^{(1)}\},\ D^{(2)} = \{d_1^{(2)}, \ldots, d_m^{(2)}\}$. Equation (2.14) can be decoupled into two problems, thus the least-squares solutions for the dictionaries are obtained by computing

$$\begin{aligned}
D^{(1)} &= (\sum_{i=1}^{N} x_i^{(1)}(\alpha_i^{(1)})^T)(\sum_{i=1}^{N} \alpha_i^{(1)}(\alpha_i^{(1)})^T)^{-1}, \\
D^{(2)} &= (\sum_{i=1}^{N} x_i^{(2)}(\alpha_i^{(2)})^T)(\sum_{i=1}^{N} \alpha_i^{(2)}(\alpha_i^{(2)})^T)^{-1}.
\end{aligned} \tag{2.15}$$

Projection onto the ℓ_2 ball is achieved by

$$d_j^{(1)} = d_j^{(1)} / \max(\|d_j^{(1)}\|_2, 1),\ d_j^{(2)} = d_j^{(2)} / \max(\|d_j^{(2)}\|_2, 1).$$

Finally by fixing $D^{(1)}$, $D^{(2)}$, $\alpha_i^{(1)}$ and $\alpha_i^{(2)}$, I can update W by solving,

$$\min_W \sum_{i=1}^{N} \gamma_1 \|\alpha_i^{(2)} - W\alpha_i^{(1)}\|_w^2 + \gamma_2 \|W\|_F^2. \tag{2.16}$$

The closed-form solution of Equation (2.16) is:

$$W = (\sum_{i=1}^{N} \gamma_1 \alpha_i^{(2)} (\alpha_i^{(1)})^T)(\sum_{i=1}^{N} \gamma_1 \alpha_i^{(1)} (\alpha_i^{(1)})^T + \gamma_2 \mathbf{I})^{-1}.$$

2.6 Conclusion

This chapter introduced methods related to the work in this dissertation. First a brief introduction to image registration was given.Then, the image analogy method was introduced. It is a simple method to relate image patches from different modalities. Then dictionary learning was explained to learn an over-complete basis to represent the data in a space. Similar to dictionary learning, coupled dictionary learning was discussed to learn an over-complete basis for the data in two coupled spaces.

CHAPTER 3: COUPLED DICTIONARY LEARNING FOR MULTI-MODAL REGISTRATION

Multi-modal registration is a crucial step to analyze images from different modalities such as correlative microscopy images (Lemoine et al., 1994; Caplan et al., 2011). The goal of image registration is to compute the spatial alignment between images by maximizing the similarity between images in the space of admissible transformations. Here, defining image similarity is challenging as images may have strikingly different appearances. Hence, standard image similarity measures may not apply. In this chapter, I propose a method to transform the appearance of an image from one modality to another. By transforming the appearance of the image, a multi-modal registration problem can be simplified to a mono-modal one. The appearance transformation can be realized by image analogy[1]. The proposed image analogy method is based on coupled dictionary learning. The flowchart of my method is shown in Figure 3.1.

3.1 Introduction

Correlative microscopy integrates different microscopy technologies including conventional light-, confocal- and electron transmission microscopy (Caplan et al., 2011) for the improved examination of biological specimens. For example, fluorescent markers can be used to highlight regions of interest combined with an electron-microscopy image to provide high-resolution structural information of the regions. To allow such joint analysis requires the registration of multi-modal microscopy images. This is a challenging problem due to (large) appearance differences between the image modalities.

[1]The standard image analogy method is introduced in Section 2.2.

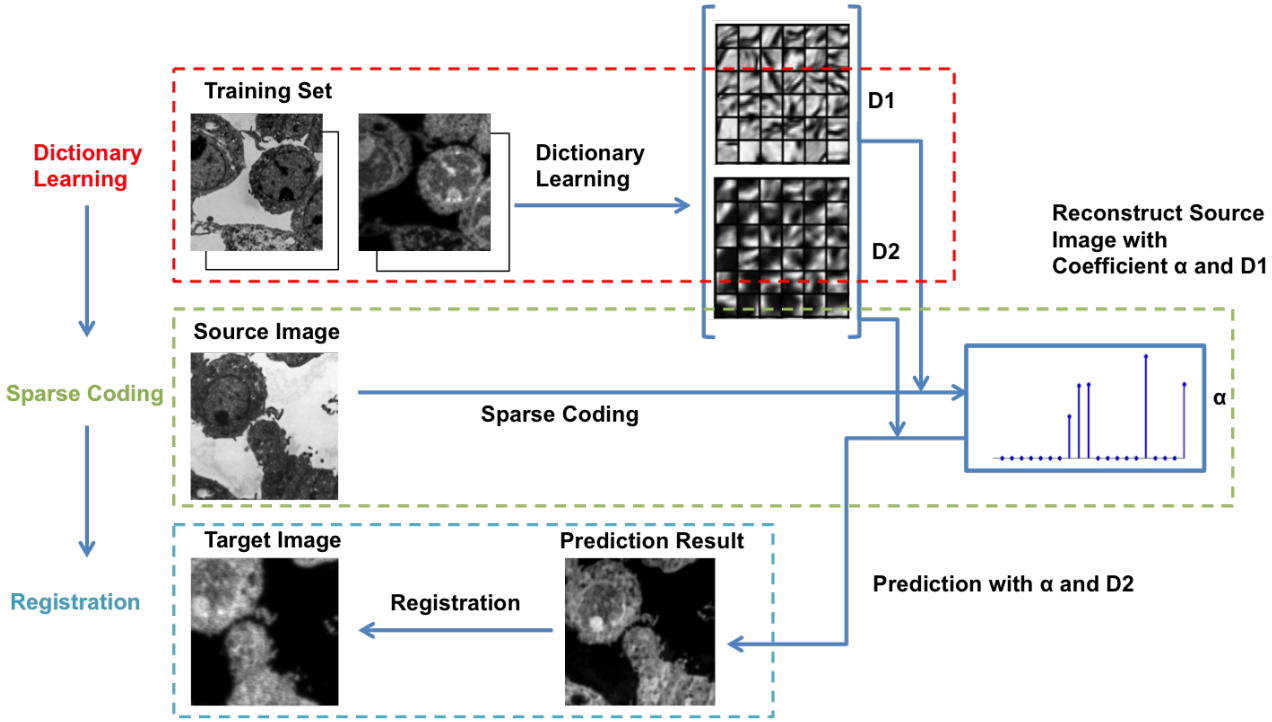


Figure 3.1: Flowchart of the proposed method. This method has three components: 1. dictionary learning: learning coupled dictionaries for both training images from different modalities; 2. sparse coding: computing sparse coefficients for the learned dictionaries to reconstruct the source image while 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.

Image registration which was introduced in Section 2.1 is an essential part of many image analysis approaches. The registration of correlative microscopic images is very challenging: images should carry distinct information to combine, for example, knowledge about protein locations (using fluorescence microscopy) and high-resolution structural data (using electron microscopy). However, this precludes the use of simple alignment measures such as the sum of squared intensity differences because intensity patterns do not correspond well or a multi-channel image has to be registered to a gray-valued image.

A solution for the registration of correlative microscopy images is to perform landmark-based alignment (Section 2.1), which can be greatly simplified by adding fiducial markers (Fronczek et al., 2011). Fiducial markers cannot easily be added to some specimens, hence an alternative intensity-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. For example, multichannel microscopy images of cells can be registered by registering their cell segmentations (Yang et al., 2008). However, such image-based approaches are highly application-specific and difficult to devise for the non-expert. If the images are structurally similar (for example when aligning EM images of different resolutions (Kaynig et al., 2007), standard feature point detectors can be used.

In this chapter I therefore propose a method inspired by early work on texture synthesis in computer graphics using image analogies (Hertzmann et al., 2001). Here, the objective is to transform the appearance of one image to the appearance of another image (for example transforming an expressionistic into an impressionistic painting). The transformation rule is learned based on example image pairs. For image registration this amounts to providing a set of (manually) aligned images of the two modalities to be registered from which an appearance transformation rule can be learned. A multi-modal registration problem can then be converted into a mono-modal one. The learned transformation rule is still highly application-specific, however it only requires manual

alignment of sets of training images which can easily be accomplished by a domain specialist who does not need to be an expert in image registration.

Arguably, transforming image appearance is not necessary if using an image similarity measure which is invariant to the observed appearance differences. In medical imaging, mutual information (MI) (Wells et al., 1996) is the similarity measure of choice for multi-modal image registration. I show for two correlative microscopy example problems that MI registration is indeed beneficial, but that registration results can be improved by combining MI with an image analogies approach. To obtain a method with better generalizability than standard image analogies (Hertzmann et al., 2001) I devise an image analogies method using ideas from sparse coding (Bruckstein et al., 2009), where corresponding image patches are represented by a learned basis (a dictionary). Dictionary elements capture correspondences between image patches from different modalities and therefore allow to transform one modality to another modality.

This chapter is organized as follows: First, I briefly introduce some related work in Section 3.2. Section 3.3 describes the proposed method for multi-modal registration. Image registration results are shown and discussed in Section 3.4. The chapter concludes with a summary of results and an outlook on future work in Section 3.5.

3.2 Related work

3.2.1 Multi-modal image registration for correlative microscopy

Since correlative microscopy combines different microscopy modalities, resolution differences between images are common. This poses challenges with respect to finding corresponding regions in the images. If the images are structurally similar (for example when aligning EM images of different resolutions (Kaynig et al., 2007)), standard feature point detectors can be used.

There are two groups of methods for more general multi-modal image registration (Wachinger and Navab, 2010). The first set of approaches applies advanced similarity measures, such as mutual information (Wells et al., 1996). The second group of techniques includes methods that transform a multi-modal to a mono-modal registration (Wein et al., 2008). For example, Wachinger introduced entropy images and Laplacian images which are general structural representations (Wachinger

and Navab, 2010). The motivation of the proposed method is similar to Wachinger's approach, i.e. transform the modality of one image to another, but I use an image analogy approach to achieve this goal thereby allowing for the reconstruction of a microscopy image in the appearance space of another.

3.2.2 Image synthesis

Image synthesis is a process to create new images from given images or image descriptions (Fisher et al., 1996). Image analogy, first introduced in (Hertzmann et al., 2001), is an image synthesis method which has been used widely in texture synthesis. In (Prince et al., 1995; Fischl et al., 2004), the authors proposed a method to synthesize magnetic resonance (MR) images. In (Roy et al., 2011), the authors introduced a compressed sensing based approach for MR tissue contrast synthesis, however, the dictionary is generated by random selection from the training data. In (van Tulder and de Bruijne, 2015), the authors proposed a restricted Boltzmann machines (RBM) based approach to synthesize MRI images for classification, however, the authors only reported the accuracy and whether the method improved the image synthesis results (how similar a synthesized image is compared to the original image is unknown). In (Iglesias et al., 2013), the authors applied a simplified image analogy method to demonstrate the benefits of synthesis in registration and segmentation for MRI images.

For multi-modal image registration, an image synthesis method can be used to transfer a given image from one modality to another. Then the multi-modal image registration problem simplifies to a mono-modal one. I use image analogy for modality transformation in this chapter because image analogy is a simple and straightforward image synthesis method and many other applications are based on this method such as image super-resolution (Freeman et al., 2002) and image completion (Drori et al., 2003).

However, since this method uses a nearest neighbor (NN) search of the image patch centered at each pixel, the resulting images are usually noisy because the ℓ_2 norm based NN search does not preserve the local consistency well (see Figure 2.3 (d)) (Hertzmann et al., 2001). This problem can be partially solved by a multi-scale search and a coherence search which enforce local consistency

among neighboring pixels, but an effective solution is still missing. I introduce a dictionary learning based image analogy method to address this problem in Section 3.3.1.

The contribution of this chapter is two-fold.

- I introduce a sparse representation model for image analogy with the goal of improving the image analogy accuracy.
- I simplify multi-modal image registration by using the image analogy approach to convert the registration problem to a mono-modal registration problem.

3.3 Method

The standard coupled dictionary learning (CDL) approach has been introduced in Section 2.5. This section introduces the CDL based image analogy method.

3.3.1 CDL for image analogies

For the registration of correlative microscopy images, given two training images A and A' from different modalities, image B can be transformed to the other modality by synthesizing B' using the image analogy approach. Consider the sparse dictionary-based image denoising/reconstruction, u, given by minimizing

$$E(u, \{\alpha_i\}) = \frac{\gamma}{2}\|Lu - f\|_2^2 + \frac{1}{2}\sum_{i=1}^{N} \|R_i u - D\alpha_i\|_V^2 + \lambda\|\alpha_i\|_1, \tag{3.1}$$

where f is the given (potentially noisy) image, D is the dictionary, $\{\alpha_i\}$ are the patch coefficients, R_i selects the ith patch from the image reconstruction u, γ, $\lambda > 0$ are balancing constants, L is a linear operator (e.g., describing a convolution), and the norm is defined as $\|x\|_v^2 = x^T V x$, where $V > 0$ is positive definite. I jointly optimize for the coefficients and the reconstructed/denoised

image. Formulation Equation (3.1) can be extended to image analogies by minimizing

$$E(u^{(1)}, u^{(2)}, \{\alpha_i\}) = \frac{\gamma}{2}\|L^{(1)}u^{(1)} - f^{(1)}\|_2^2 + \frac{1}{2}\sum_{i=1}^{N}\|R_i \begin{pmatrix} u^{(1)} \\ u^{(2)} \end{pmatrix} - \begin{pmatrix} D^{(1)} \\ D^{(2)} \end{pmatrix} \alpha_i\|_V^2 + \lambda\|\alpha_i\|_1, \tag{3.2}$$

where I have corresponding 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 $f^{(2)}$). 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.

Patch-based (non-sparse) denoising has also been proposed for the denoising of fluorescence microscopy images (Boulanger et al., 2010). A conceptually similar approach using sparse coding and image patch transfer has been proposed to relate different magnetic resonance images in (Roy et al., 2011). However, this approach does not address dictionary learning or spatial consistency considered in the sparse coding stage. My approach addresses both and learns the dictionaries $D^{(1)}$ and $D^{(2)}$ explicitly.

3.3.2 Numerical solution

To simplify the optimization process of Equation (3.2), I apply an alternating optimization approach (Elad, 2010) which initializes $u^{(1)} = f^{(1)}$ and $u^{(2)} = D^{(2)}\alpha$ at the beginning, and then computes the optimal α (the dictionaries $D^{(1)}$ and $D^{(2)}$ are assumed known here). Thus the minimization problem breaks into many smaller subparts, for each subproblem I have,

$$\hat{\alpha} = \arg\min_{\alpha} \frac{1}{2}\|R_i u^{(1)} - D^{(1)}\alpha_i\|_2^2 + \lambda\|\alpha_i\|_1, \quad i \in 1, ...N. \tag{3.3}$$

Following (Li and Osher, 2009) I use a coordinate descent algorithm to solve Equation (3.3).

Given $A = D^{(1)}$, $x = \alpha_i$ and $b = R_i u^{(1)}$, then Equation (3.3) can be rewritten in the general form

$$\hat{x} = \arg\min_x \frac{1}{2}\|b - Ax\|_2^2 + \lambda\|x\|_1. \tag{3.4}$$

The coordinate descent algorithm to solve Equation (3.4) is described in Algorithm 2. This algorithm minimizes Equation (3.4) with respect to one component of x in one step, keeping all other components constant. This step is repeated until convergence.

After solving Equation (3.3), I can fix α and then update $u^{(1)}$. Now the optimization of Equation (3.2) can be changed to

$$\hat{u}^{(1)} = \arg\min_{u^{(1)}} \frac{\gamma}{2}\|u^{(1)} - f^{(1)}\|_2^2 + \sum_{i=1}^{N} \frac{1}{2}\|R_i u^{(1)} - D^{(1)}\alpha_i\|_2^2. \tag{3.5}$$

The closed-form solution of Equation (3.5) is as follows[2],

$$\hat{u}^{(1)} = (\gamma I + \sum_{i=1}^{N} R_i^T R_i)^{-1}(\gamma f^{(1)} + \sum_{i=1}^{N} R_i^T D^{(1)}\alpha_i). \tag{3.6}$$

I iterate the optimization with respect to $u^{(1)}$ and α to convergence. Then $u^{(2)} = (\sum_i^N R_i^T R_i)^{-1} D^{(2)}\hat{\alpha}$.

3.3.3 Intensity normalization

The image analogy approach may not be able to achieve a perfect prediction because: (i) image intensities are normalized and hence the original dynamic range of the images is not preserved and (ii) image contrast may be lost as the reconstruction is based on the weighted averaging of patches. To reduce the intensity distribution discrepancy between the predicted image and original image, in this method, I apply intensity normalization (normalize the different dynamic ranges of different images to the same scale for example $[0, 1]$) to the training images before dictionary learning, and also to the image analogy results.

[2]Refer to appendix 1 for more details.

3.3.4 Use in multi-modal image registration

For image registration, I (i) reconstruct the "missing" analogous image and (ii) consistently denoise the given image to be registered with (Elad and Aharon, 2006). 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. I consider rigid followed by affine and B-spline registrations in this chapter and use the implementation of the Elastix toolbox (Klein et al., 2010; Ibanez et al., 2003). As similarity measures I use sum of squared differences (SSD) and mutual information (MI). A standard gradient descent is used for optimization. For B-spline registration, I use displacement magnitude regularization which penalizes $\|T(x) - x\|^2$, where $T(x)$ is the transformation of coordinate x in an image (Klein et al., 2010). This is justified as I do not expect large deformations between the images as they represent the same structure. Hence, small displacements are expected, which are favored by this form of regularization.

3.4 Results

3.4.1 Data

I use both 2D correlative SEM/confocal images with fiducials and TEM/confocal images of mouse brains for the experiment. All experiments are performed on a Dell OptiPlex 980 computer with an Intel Core i7 860 2.9GHz CPU. The data description is shown in Table 3.1.

Table 3.1: Data Description

	Data Types	
	SEM/confocal	TEM/confocal
Number of datasets	8	6
Fiducial	100 nm gold	none
Pixel Size	40 nm	0.069μm

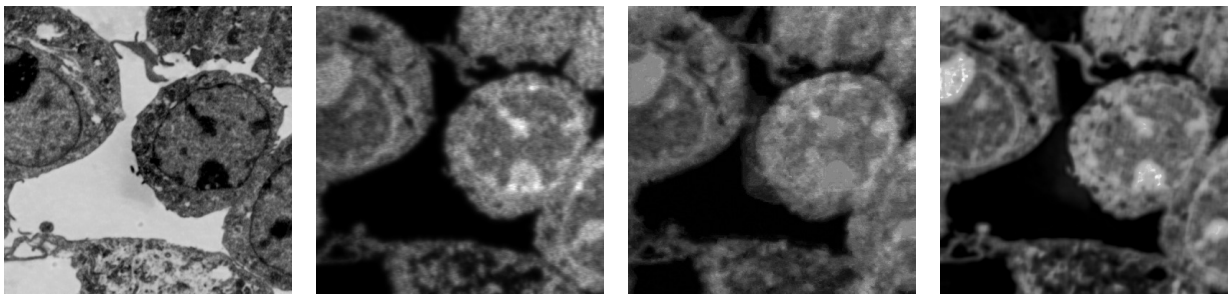
3.4.2 Registration of SEM/confocal images (with fiducials)

3.4.2.1 Data preparation

The confocal images are denoised by the sparse representation based denoising method (Elad, 2010). I use a landmark based registration on the fiducials to obtain the gold standard alignment results. The image size is about 400×400 pixels.

3.4.2.2 Image analogy (IA) results

I apply the standard IA method and the proposed method. I train the dictionaries using a leave-one-out approach. The training image patches are extracted from pre-registered SEM/confocal images as part of the preprocessing described in Section 3.4.2.1. In both IA methods I use 10×10 patches, and in the proposed method I randomly sample 50000 patches and learn 1000 dictionary elements in the dictionary learning phase. The learned dictionaries are shown in Figure 3.4. I choose $\gamma = 1$ and $\lambda = 0.15$ in Equation (3.2). In Figure 3.2, both IA methods can reconstruct the confocal image very well but the proposed method preserves more structure than the standard IA method. I also show the prediction errors and the statistical scores of the proposed IA method and standard IA method for SEM/confocal images in Table 3.2. The prediction error is defined as the sum of squared intensity differences between the predicted confocal image and the original confocal image. My method is based on patch-by-patch prediction using the learned multi-modal dictionary. Given a particular patch-size the number of sparse coding problems in the model changes linearly with the number of pixels in an image. My method is much faster than the standard image analogies method which involves an exhaustive search of the whole training set as my method is based on a dictionary representation. For example, my method takes about 500 secs for a 1024×1024 image with image patch size 10×10 and dictionary size 1000 while the standard image analogy method takes more than 30 mins for the same patch size. The CPU processing time for SEM/confocal data is shown in Table 3.3. I also illustrate the convergence of solving Equation (2.6) for both SEM/confocal and TEM/confocal images in Figure 3.6 which shows that 100 iterations are sufficient for both datasets. The IA results are shown in Figure 3.2.



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

Figure 3.2: Results of estimating a confocal (b) from an SEM image (a) using the standard IA (c) and the proposed IA method (d).

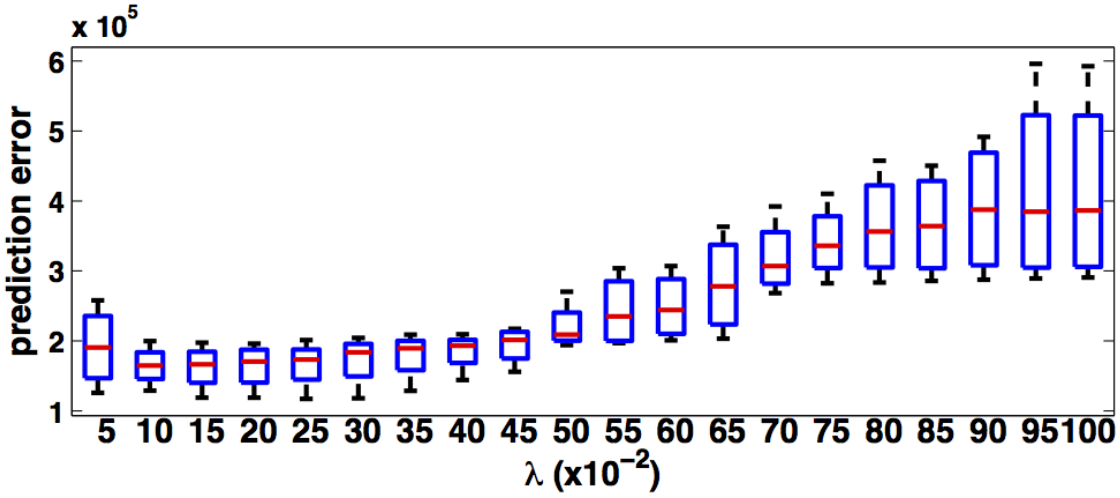


Figure 3.3: Prediction errors with respect to different λ values for SEM/confocal image. The λ values are tested from 0.05-1.0 with step size 0.05.

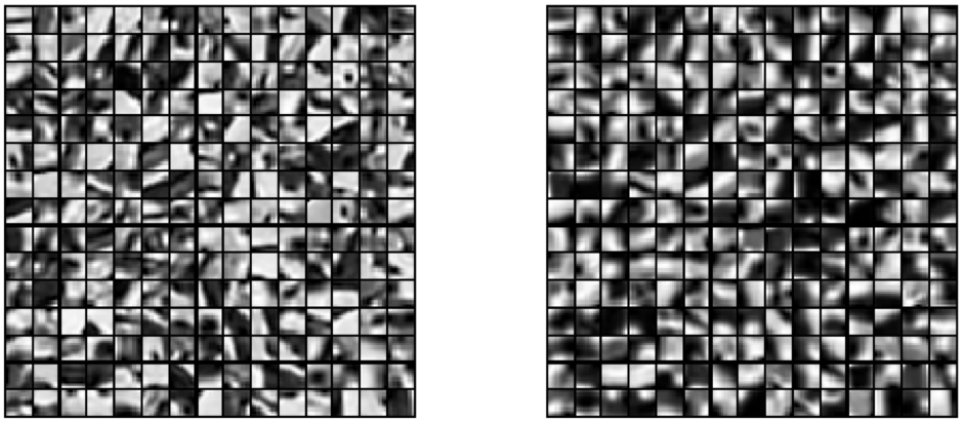


Figure 3.4: Results of dictionary learning: the left dictionary is learned from the SEM and the corresponding right dictionary is learned from the confocal image.

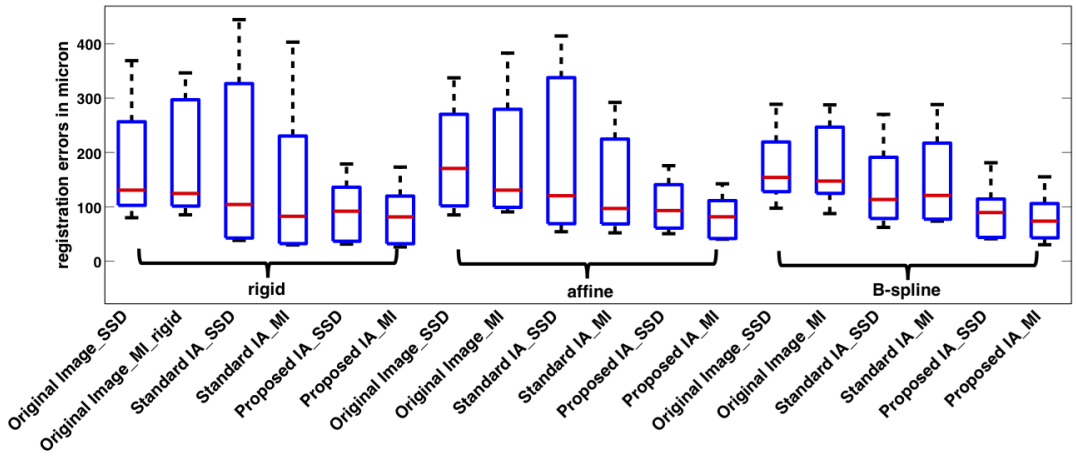


Figure 3.5: Box plot for the registration results of SEM/confocal images on landmark errors of different methods with three transformation models: rigid, affine and B-spline. The registration methods include: Original Image_SSD and Original Image_MI, registrations with original images based on SSD and MI metrics respectively; Standard IA_SSD and Standard IA_MI, registration with standard IA algorithm based on SSD and MI metrics respectively; Proposed IA_SSD and Proposed IA_MI, registration with the proposed IA algorithm based on SSD and MI metrics respectively. The bottom and top edges of the boxes are the 25th and 75th percentiles, the central red lines are the medians.

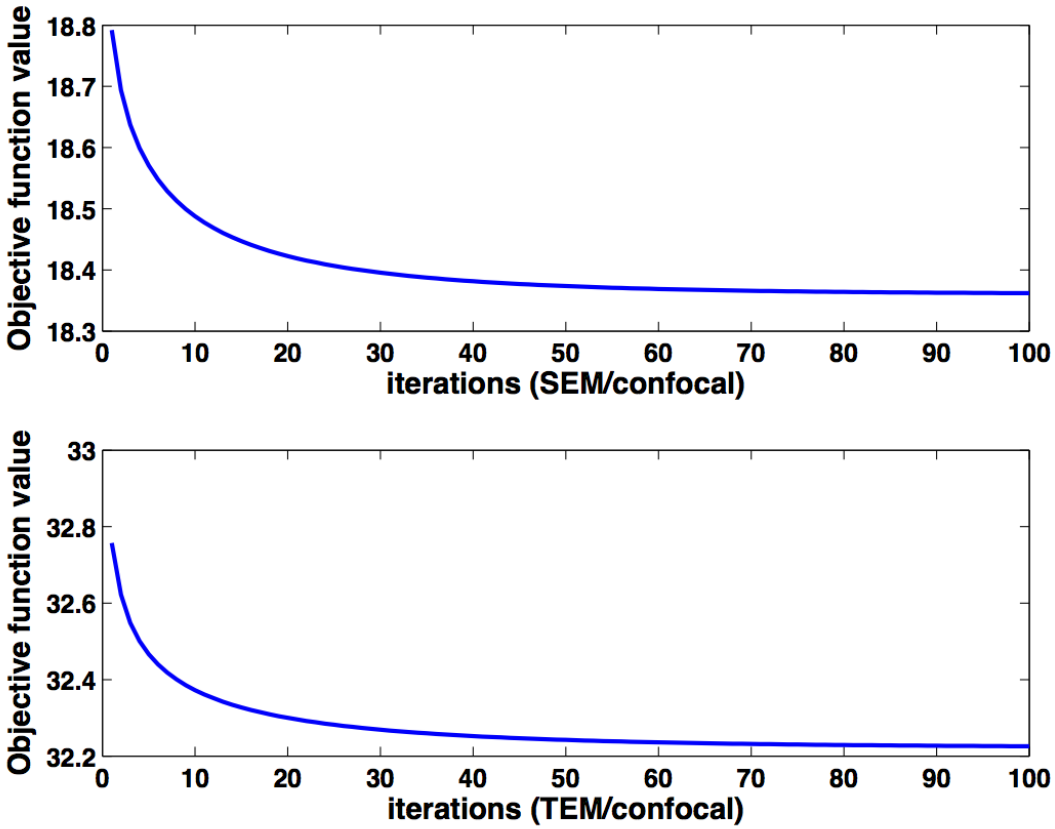


Figure 3.6: Convergence test on SEM/confocal and TEM/confocal images. The objective function is defined as in Equation (2.11). The maximum iteration number is 100. The patch size for SEM/confocal images and TEM/confocal images are 10×10 and 15×15 respectively.

Table 3.2: Prediction results for SEM/confocal images. Prediction is based on the proposed IA and standard IA methods. I used 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	$\mathbf{0.0002}$
Standard IA	2.83×10^5	7.11×10^4	

3.4.2.3 Image registration results

I resampled the estimated confocal images with up to $\pm 600\ nm$ (15 pixels) in translation in the x and y directions (at steps of 5 pixel) and $\pm 15^\circ$ in rotation (at steps of 5 degree) with respect to the gold standard alignment. Then I registered the resampled estimated confocal images to the corresponding original confocal images. The goal of this experiment is to test the ability of the proposed methods to recover from misalignments by translating and rotating the pre-aligned image within a practically reasonable range. Such a rough initial automatic alignment can for example be achieved by image correlation. The image registration results based on both image analogy methods are compared to registration results using original images using both SSD and MI as similarity measures[3]. Table 3.4 summarizes the registration results on translation and rotation errors based on the rigid transformation model for each image pair over all these experiments. The results are reported as physical distances instead of pixels. I also perform registrations using affine and B-spline transformation models. These registrations are initialized with the result from the rigid registration. Figure 3.5 shows the box plot for all the registration results. The proposed method achieves the best registration results for all three transformation models compared with directly registering the multi-modal images and the standard image analogy method. However, the proposed methods increase the registration results little with more flexible transformation models i.e. affine and b-spline models.

[3]I invert the grayscale values of the original SEM image for SSD based image registration of the SEM/confocal images.

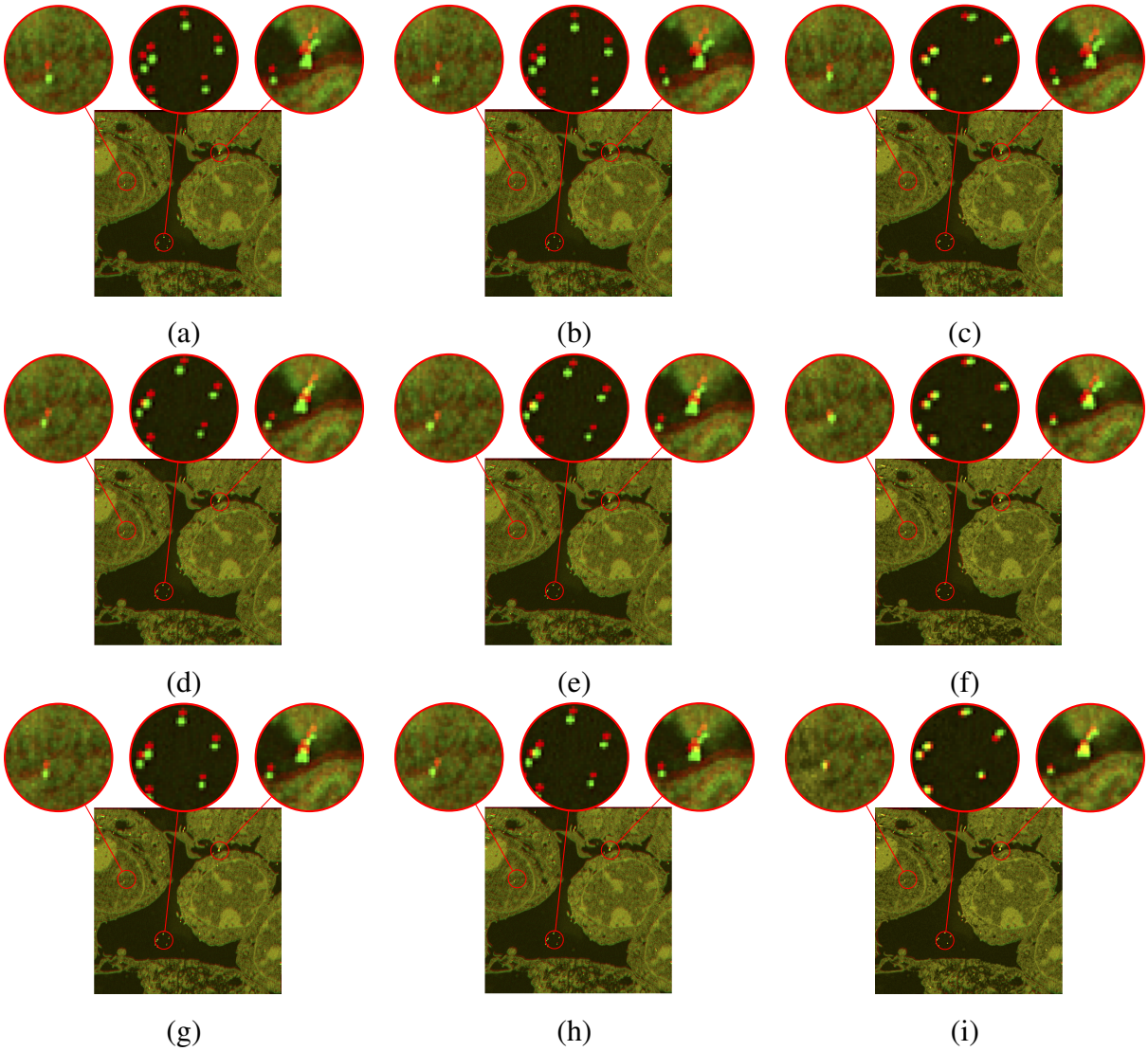


Figure 3.7: Results of registration for SEM/confocal images using MI similarity measure with direct registration (first row), standard IA (second row) and the proposed IA method (third row) for (a,d,g) rigid registration (b,e,h) affine registration and (c,f,i) b-spline registration. Some regions are 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. Differences are generally very small indicating that for these images a rigid transformation model may already be sufficiently good.

Table 3.3: CPU time (in seconds) for SEM/confocal images. The p-value is computed using a paired t-test.

Method	mean	std	p-value
Proposed IA	**82.2**	6.7	**0.00006**
Standard IA	407.3	10.1	

3.4.2.4 Hypothesis test on registration results

In order to check whether the registration results from different methods are statistically different from each other, I use hypothesis testing (Weiss and Weiss, 2012). I assume the registration results (rotations and translations) are independent and normally distributed random variables with means μ_i and variances σ_i^2. For the results from 2 different methods, the null hypothesis (H_0) is $\mu_1 = \mu_2$, and the alternative hypothesis (H_1) is $\mu_1 \leq \mu_2$. I apply the one-sided paired sample t-test for equal means using MATLAB (MATLAB, 2012). The level of significance is set at 5%. Based on the hypothesis test results in Table 3.5, the proposed method shows significant differences with respect to the standard IA method for the registration error with respect to the SSD metric on rigid registration and both MI and SSD metrics for affine and B-spline registrations. Table 3.4 and Table 3.5 also show that the proposed method outperforms the standard image analogy method as well as the direct use of mutual information on the original images in terms of registration accuracy. However, as deformations are generally relatively rigid no statistically significant improvements in registration results could be found within a given method relative to the different transformation models as illustrated in Table 3.6.

Table 3.4: SEM/confocal rigid registration errors on translation (t) and rotation (r)($t = \sqrt{t_x^2 + t_y^2}$ where t_x and t_y are translation errors in x and y directions respectively; t is in nm; pixel size is $40nm$; r is in degree.) Here, the registration methods include: Original Image_SSD and Original Image_MI, registrations with original images based on SSD and MI metrics respectively; Standard IA_SSD and Standard IA_MI, registration with standard IA algorithm based on SSD and MI metrics respectively; Proposed IA_SSD and Proposed IA_MI, registration with the proposed IA algorithm based on SSD and MI metrics respectively.

	r_{mean}	r_{median}	r_{std}	t_{mean}	t_{median}	t_{std}
Proposed IA_SSD	0.357	0.255	0.226	92.457	91.940	56.178
Proposed IA_MI	**0.239**	0.227	0.077	**83.299**	81.526	54.310
Standard IA_SSD	0.377	0.319	0.215	178.782	104.362	162.266
Standard IA_MI	0.3396	0.332	0.133	140.401	82.667	135.065
Original Image_SSD	0.635	0.571	0.215	170.484	110.317	117.371
Original Image_MI	0.516	0.423	0.320	176.203	104.574	116.290

Table 3.5: Hypothesis test results (p-values) with multiple testing correction results (FDR corrected p-values in parentheses) for registration results evaluated via landmark errors for SEM/confocal images. I use a one-sided paired t-test. Comparison of different image types (original image, standard IA, proposed IA) using the same registration models (rigid, affine, B-spline). The proposed model shows the best performance for all transformation models. (**Bold** indicates statistically significant improvement at significance level $\alpha = 0.05$ after correcting for multiple comparisons with FDR (Benjamini and Hochberg, 1995).)

		Original Image/Standard IA	Original Image/Proposed IA	Standard IA/Proposed IA
Rigid	SSD	0.5017 (0.5236)	**0.0040** (0.0102)	0.0482 (0.0668)
	MI	0.0747 (0.0961)	**0.0014** (0.0052)	0.0888(0.1065)
Affine	SSD	0.5236 (0.5236)	**0.0013** (0.0052)	0.0357 (0.0535)
	MI	**0.0298** (0.0488)	**0.0048** (0.0108)	**0.0258** (0.0465)
B-spline	SSD	**0.0017** (0.0052)	**0.0001** (0.0023)	**0.0089** (0.0179)
	MI	0.1491 (0.1678)	**0.0002** (0.0024)	**0.0017** (0.0052)

Table 3.6: Hypothesis test results (p-values) with multiple testing correction results (FDR corrected p-values in parentheses) for registration results measured via landmark errors for SEM/confocal images. I use a one-sided paired t-test. Comparison of different registration models (rigid, affine, B-spline) within the same image types (original image, standard IA, proposed IA). Results are not statistically significantly better after correcting for multiple comparisons with FDR.)

		Rigid/Affine	Rigid/B-spline	Affine/B-spline
Original Image	SSD	0.7918 (0.8908)	0.3974 (0.6596)	0.1631 (0.5873)
	MI	0.6122 (0.7952)	0.3902 (0.6596)	0.3635 (0.6596)
Standard IA	SSD	0.9181 (0.9371)	0.1593 (0.5873)	0.0726 (0.5873)
	MI	0.5043 (0.7564)	0.6185 (0.7952)	0.7459 (0.8908)
Proposed IA	SSD	0.9371 (0.9371)	0.3742 (0.6596)	0.0448 (0.5873)
	MI	0.4031 (0.6596)	0.1616 (0.5873)	0.2726 (0.6596)

3.4.2.5 Discussion

From Figure 3.5, the improvement of registration results within an individual registration model from rigid registration to affine and B-spline registrations are not significant due to the fact that both SEM/confocal images are acquired from the same piece of tissue section. The rigid transformation model can capture the deformation well enough, though small improvements can visually be observed using more flexible transformation models as illustrated in the composition images between the registered SEM images using three registration methods (direct registration and the two IA methods) and the registered SEM images based on fiducials of Figure 3.7. The proposed method can achieve the best results for all the three registration models. See also Table 3.6. The results show that the similarity measures matters (my method is best) for the registration. However, the different transformation models do not matter much for this experiment.

3.4.3 Registration of TEM/confocal images (without fiducials)

3.4.3.1 Data preparation

The corresponding confocal image and TEM image are resampled to an intermediate resolution. The final resolution is 14.52 pixels per μm, and the image size is about 256×256 pixels. The datasets are already roughly registered based on manually labeled landmarks with a rigid transformation model.

3.4.3.2 Image analogy results

I tested the standard image analogy method and our proposed sparse method. For both image analogy methods I use 15×15 patches, and for the proposed method I randomly sample 50000 patches and learn 1000 dictionary elements in the dictionary learning phase. The learned dictionaries are shown in Figure 3.9. I choose $\gamma = 1$ and $\lambda = 0.1$ in Equation (3.2). The image analogies results in Figure 3.10 show that the proposed method preserves more local structure than the standard image analogy method. I show the prediction error of the proposed IA method and standard IA method for TEM/confocal images in Table 3.7. The CPU processing time for the TEM/confocal data is given in Table 3.8.

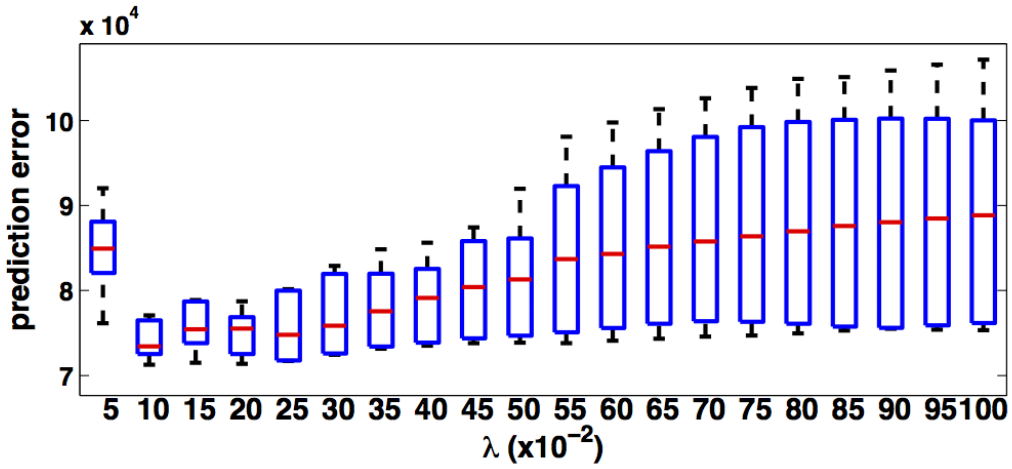


Figure 3.8: Prediction errors for different λ values for TEM/confocal image. The λ values are tested from 0.05-1.0 with step size 0.05.

3.4.3.3 Image registration results

I manually determined $10 \sim 15$ corresponding landmark pairs on each dataset to establish a gold standard for registration. The same type and magnitude of shifts and rotations as for the SEM/confocal experiment are applied. The image registration results based on both image analogy methods are compared to the landmark based image registration results using mean absolute errors (MAE) and standard deviations (STD) of the absolute errors on all the corresponding landmarks. I use both SSD and mutual information (MI) as similarity measures. The registration results are

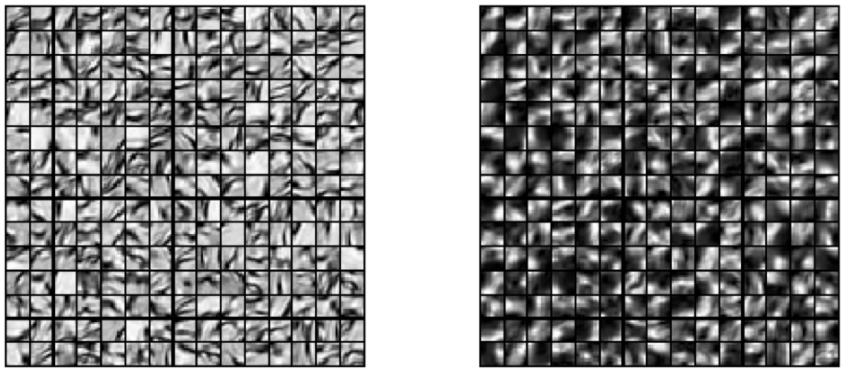
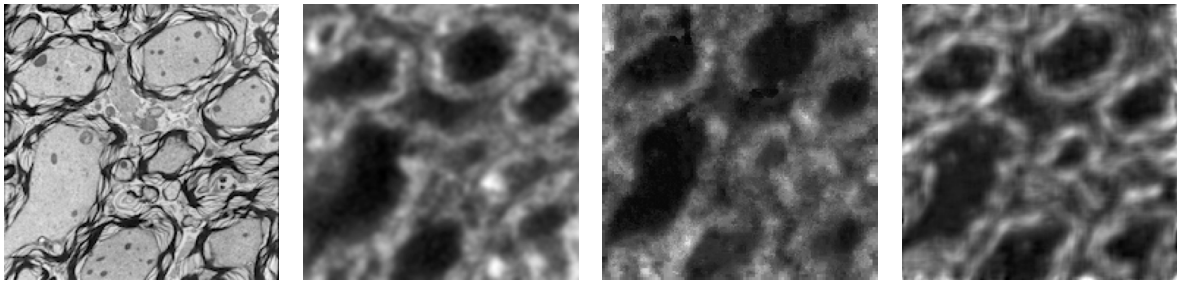


Figure 3.9: Results of dictionary learning: the left dictionary is learned from the TEM and the corresponding right dictionary is learned from the confocal image.



(a) TEM image (b) Confocal image (c) Standard IA (d) Proposed IA

Figure 3.10: Result of estimating the confocal image (b) from the TEM image (a) for the standard image analogy method (c) and the proposed sparse image analogy method (d) which shows better preservation of structure.

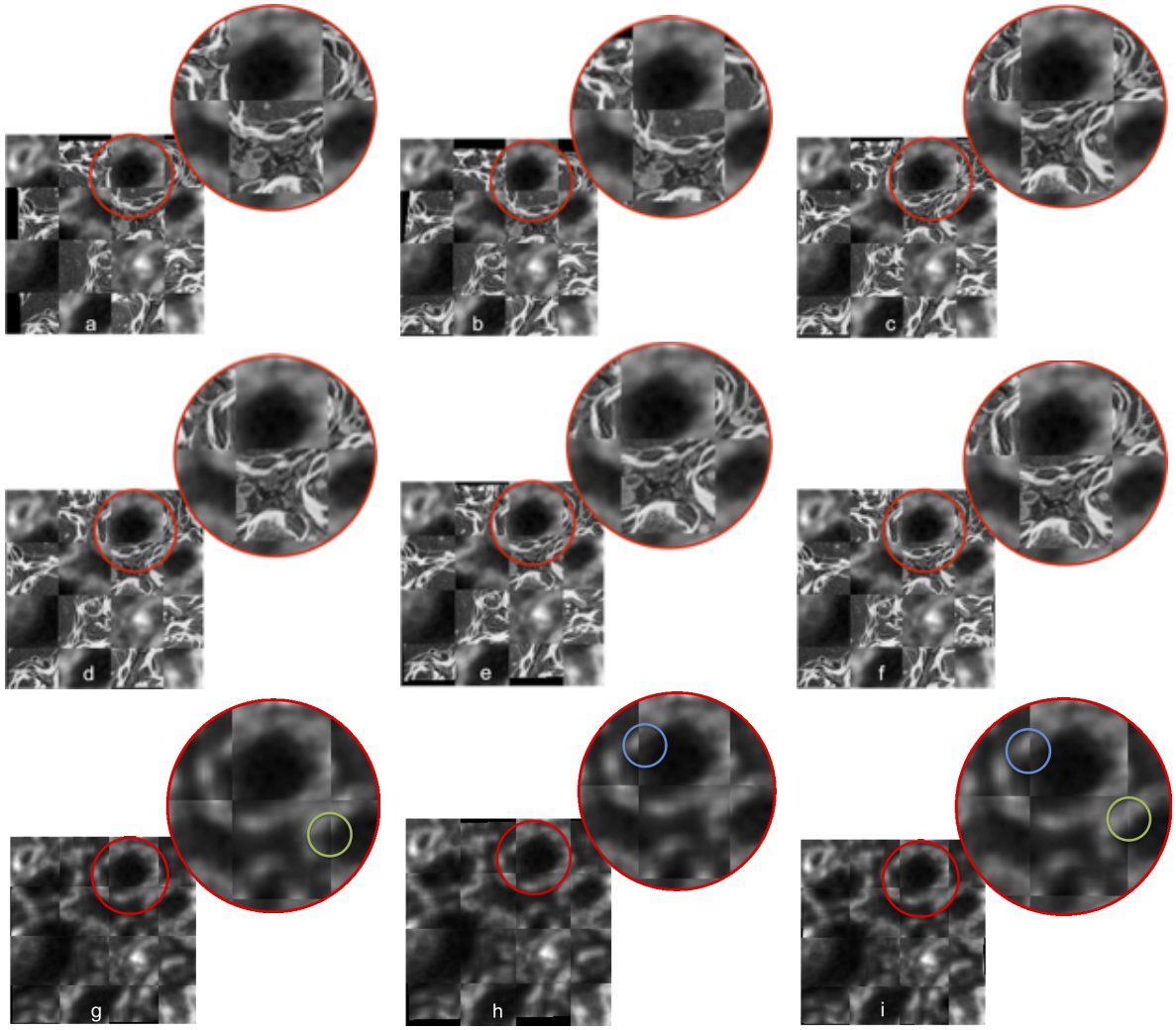


Figure 3.11: Results of registration for TEM/confocal images using MI similarity measure with directly registration (first row) and the proposed IA method (second and third rows) using (a,d,g) rigid registration (b,e,h) affine registration and (c,f,i) b-spline registration. The results are shown in a checkerboard image for comparison. Here, first and second rows show the checkerboard images of the original TEM/confocal images while the third row shows the checkerboard image of the results of the proposed IA method. Differences are generally small, but some improvements can be observed for B-spline registration. The grayscale values of the original TEM image are inverted for better visualization.

Table 3.7: Prediction results for TEM/confocal images. Prediction is based on the proposed IA and standard IA methods, and I use SSR to evaluate the prediction results. The p-value is computed using a paired t-test.

Method	mean	std	p-value
Proposed IA	$\mathbf{7.43 \times 10^4}$	4.72×10^3	$\mathbf{0.0015}$
Standard IA	8.62×10^4	6.37×10^3	

Table 3.8: CPU time (in seconds) for TEM/confocal images. The p-value is computed using a paired t-test.

Method	mean	std	p-value
Proposed IA	$\mathbf{35.2}$	4.4	$\mathbf{0.00019}$
Standard IA	196.4	8.1	

displayed in Figure 3.12 and Table 3.9. The landmark based image registration result is the best result achievable given the affine transformation model. I show the results for both image analogy methods as well as using the original TEM/confocal image pairs[4]. Figure 3.12 shows that the MI based image registration results are similar among the three methods and the proposed method performs slightly better. The results are reported as physical distances instead of pixels. Also the results of the proposed method are close to the landmark based registration results (best registration results). For SSD based image registration, the proposed method is more consistent than the other two methods for the current dataset.

3.4.3.4 Hypothesis test on registration results

I use the same hypothesis test method as in Section 3.4.2.4, and test the means of different methods on MAE of corresponding landmarks. Table 3.10 and Table 3.11 indicate that the registration result of the proposed method shows significant improvement over the result using original images with both SSD and MI metric. Also, the result of the proposed method is significantly better than the standard IA method with MI metric.

[4]I inverted the grayscale values of original TEM image for SSD based registration of original TEM/confocal images.

Table 3.9: TEM/confocal rigid registration results (in μm, pixel size is 0.069 μm).

		Proposed IA		Standard IA		Original Image		Landmark	
case		MAE	STD	MAE	STD	MAE	STD	MAE	STD
1	SSD	**0.3174**	0.2698	0.3219	0.2622	0.4352	0.2519	0.2705	0.1835
	MI	0.3146	0.2657	**0.3132**	0.2601	0.5161	0.2270		
2	SSD	**0.4911**	0.1642	0.5759	0.2160	2.5420	1.6877	0.3091	0.1594
	MI	**0.4473**	0.1869	0.4747	0.3567	1.4245	0.1780		
3	SSD	0.5379	0.2291	1.8940	1.0447	**0.5067**	0.2318	0.3636	0.1746
	MI	**0.3864**	0.2649	0.5261	0.2008	0.4078	0.2608		
4	SSD	**0.4451**	0.2194	0.4516	0.2215	0.4671	0.2484	0.3823	0.2049
	MI	0.4554	0.2298	**0.4250**	0.2408	0.4740	0.2374		
5	SSD	**0.6268**	0.2505	1.2724	0.6734	1.3174	0.3899	0.2898	0.2008
	MI	**0.3843**	0.2346	0.6172	0.2429	0.7018	0.2519		
6	SSD	**0.7832**	0.5575	0.8159	0.4975	2.2080	1.4228	0.3643	0.1435
	MI	**0.7259**	0.4809	1.2772	0.4285	0.8383	0.4430		

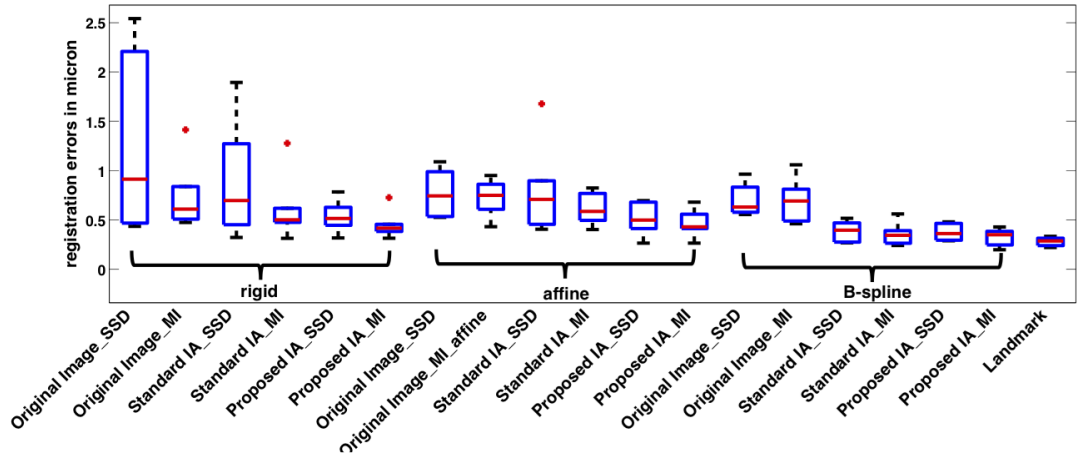


Figure 3.12: Box plot for the registration results of TEM/confocal images for different methods. The bottom and top edges of the boxes are 25th and 75th percentiles, the central red lines indicate the medians.

Table 3.10: Hypothesis test results (p-values) with multiple testing correction results (FDR corrected p-values in parentheses) for registration results evaluated via landmark errors for TEM/confocal images. I use a one-sided paired t-test. Comparison of different image types (original image, standard IA, proposed IA) using the same registration models (rigid, affine, B-spline). The proposed image analogy method performs better for affine and B-spline deformation models. (**Bold** indicates statistically significant improvement at a significance level $\alpha = 0.05$ after correcting for multiple comparisons with FDR.)

		Original Image/Standard IA	Original Image/Proposed IA	Standard IA/Proposed IA
Rigid	SSD	0.2458 (0.2919)	0.0488 (0.1069)	0.0869 (0.1303)
	MI	0.2594 (0.2919)	0.0478 (0.1069)	0.0594 (0.1069)
Affine	SSD	0.5864 (0.5864)	**0.0148** (0.0445)	0.0750 (0.1226)
	MI	0.1593 (0.2048)	**0.0137** (0.0445)	0.0556 (0.1069)
B-spline	SSD	**0.0083** (0.0445)	**0.0085** (0.0445)	0.3597 (0.3809)
	MI	0.0148 (0.0445)	**0.0054** (0.0445)	0.1164 (0.1611)

Table 3.11: Hypothesis test results (p-values) with multiple testing correction results (FDR corrected p-values in parentheses) for registration results evaluated via landmark errors for TEM/confocal images. I use a one-sided paired t-test. Comparison of different image types (original image, standard IA, proposed IA) using the same registration models (rigid, affine, B-spline). Results are overall suggestive of the benefit of B-spline registration, but except for the standard IA do not reach significance after correction for multiple comparisons. This may be due to the limited sample size. (**Bold** indicates statistically significant improvement after correcting for multiple comparisons with FDR.)

		Rigid/Affine	Rigid/B-spline	Affine/B-spline
Original Image	SSD	0.0792(0.1583)	0.1149(0.2069)	0.3058(0.4865)
	MI	0.4325(0.4865)	0.4091(0.4865)	0.3996(0.4865)
Standard IA	SSD	0.3818(0.4865)	0.0289(0.1041)	0.0280(0.1041)
	MI	0.4899(0.5188)	0.0742(0.1583)	**0.0009**(0.0177)
Proposed IA	SSD	0.3823(0.4865)	0.0365(0.1096)	0.0216(0.1041)
	MI	0.5431(0.5431)	0.0595(0.1531)	0.0150(0.1041)

3.4.3.5 Discussion

In Figure 3.12, the affine and B-spline registrations using the proposed IA method show significant improvement compared with affine and B-spline registrations on the original images. In comparison to the SEM/confocal experiment (Figure 3.7) the checkerboard image shown in Figure 3.11 shows slightly stronger deformations for the more flexible B-spline model leading to slightly better local alignment. The proposed method still achieves the best results for the three registration models. The results show that both similarity measure and transformation model matter for the data, and our proposed method combined with the B-spline model can achieve the best results.

3.5 Conclusion

I developed a multi-modal registration method for correlative microscopy. The method is based on image analogies with a sparse representation model. The proposed method can be regarded as learning a mapping between image modalities such that either an SSD or MI image similarity measure becomes appropriate for image registration. Any desired image registration model could be combined with the proposed method as long as it supports either SSD or a MI as an image similarity measure. The proposed method then becomes an image pre-processing step. I tested the

method on SEM/confocal and TEM/confocal image pairs with rigid registration followed by affine and B-spline registrations. The image registration results from Figure 3.5 and Figure 3.12 suggest that the sparse image analogy method can improve registration robustness and accuracy. While the proposed method does not show improvements for every individual dataset, the proposed method improved registration results significantly for the SEM/confocal experiments for all transformation models and for the TEM/confocal experiments for affine and B-spline registration. Furthermore, when using the proposed image analogy method multi-modal registration based on SSD becomes feasible. I also compare the runtime between the standard IA and the proposed IA methods. The proposed method runs about 5 times faster than the standard method. While the runtime is far from real-time performance, the method is sufficiently fast for correlative microscopy applications.

Future work includes additional validation on a larger number of datasets from different modalities. The goal is also to estimate the local quality of the image analogy result. This quality estimate could then be used to weight the registration similarity metrics to focus on regions of high confidence. Other similarity measures can be modulated similarly. I will also apply the sparse image analogy method to 3D images.

CHAPTER 4: SEMI-COUPLED DICTIONARY LEARNING FOR DEFORMATION ESTIMATION

Some applications require registering a common reference, atlas, image to a set of subject images, such as atlas based segmentation (Rohlfing and Maurer, 2005; Aljabar et al., 2009). The standard approach for these applications is doing pairwise image registration. In this chapter, rather than directly estimating deformations by standard image registration methods, I investigate how to obtain a basis of the space of deformations. In particular, I explore how image appearance differences with respect to a common reference image (atlas) can be used to estimate deformations represented by such a basis. I use a coupled dictionary learning method to jointly learn a basis for image appearance differences and their related deformations. The proposed method is based on local image patches. The proposed method is evaluated on synthetically generated datasets as well as on a structural magnetic resonance (MR) brain imaging dataset. The proposed method results in an improved deformation estimation accuracy while reducing the search space compared to nearest neighbor search and demonstrates that learning a deformation basis is feasible.

4.1 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. (Chou et al., 2013) and Wang et al. (Wang et al., 2013) propose prediction based models to estimate deformations based on image appearances. In (Chou et al., 2013; Kim et al., 2012), the authors propose to learn a global correlation between difference image and the deformation or the parameters for

deformation. Image intensity differences or image level features are used to learn a regression model on the corresponding deformation parameters. For a new test image, the regression models are applied to predict the deformation field. In (Shi et al., 2012; Wang et al., 2013), the authors predict the deformations on a couple of key points in a test image from a sparse linear combination of deformations from training images based on the similarity of appearances in (Shi et al., 2012), and then a dense deformation field is interpolated by the Thin-Plate Spline model in (Wang et al., 2013). However, since these methods are based on finding the correspondences between the key points in test image and training images, and key point detection is a nontrivial task (Zheng et al., 2010; Guerrero et al., 2012), I propose a more general framework to solve the problem, i.e. estimation of deformation based on difference images.

The deformation prediction problem is challenging as it is difficult to model the relationship between difference images and deformations, for example, the correlation between appearances and local deformations could be highly nonlinear. Also different deformation representation methods further complicate the modeling of the relationship. My proposed coupled dictionary learning method addresses these issues: it learns bases for both difference images 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 intensity differences relate to deformations.

The main contributions of this chapter include:

- a coupled dictionary based framework for deformation prediction using appearance;
- an illustration of the broad applicability of the method to different types of deformations;
- a demonstration that a basis for deformations can indeed be learned with a relatively small number of dictionary atoms.

This chapter is organized as follows: Section 4.2 describes the coupled dictionary learning method. Section 4.3 introduces parametrization methods for deformations. Section 4.4 applies my model to both synthetic and real data. The chapter concludes with a summary of results and an outlook on future work in Section 4.5.

4.2 Method

Image registration has been introduced in Section 2.1. There are several ways to parameterize the deformation, ϕ. I will discuss some of the most common parametrization methods in Section 4.3. The basic framework of my algorithm is illustrated in Figure 4.1. In the proposed framework, I first extract patches from pairs of training images, original and deformed, and corresponding deformation mapping, 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 images in the atlas space. Given a test image, I reconstruct its intensity differences with respect to the atlas image patch by patch with the dictionary corresponding to the image at the same time predicting the deformation with the dictionary corresponding to the deformation. The overall predicted deformation is reconstructed from local patches. The overlapping patches are averaged during reconstruction.

4.2.1 SCDL for deformations and appearances

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

4.2.2 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 $I = T - S$ (S is an input source image, and T is the common target/atlas image), similar to Equation (2.12), I solve the following sparse coding problem,

$$\{\alpha_i^{(1)}\} = \underset{\alpha_i^{(1)}}{\operatorname{argmin}} \frac{1}{2}\|I_i^{(1)} - D^{(1)}\alpha_i^{(1)}\|_2^2 + \lambda_1\|\alpha_i^{(1)}\|_1, \tag{4.1}$$

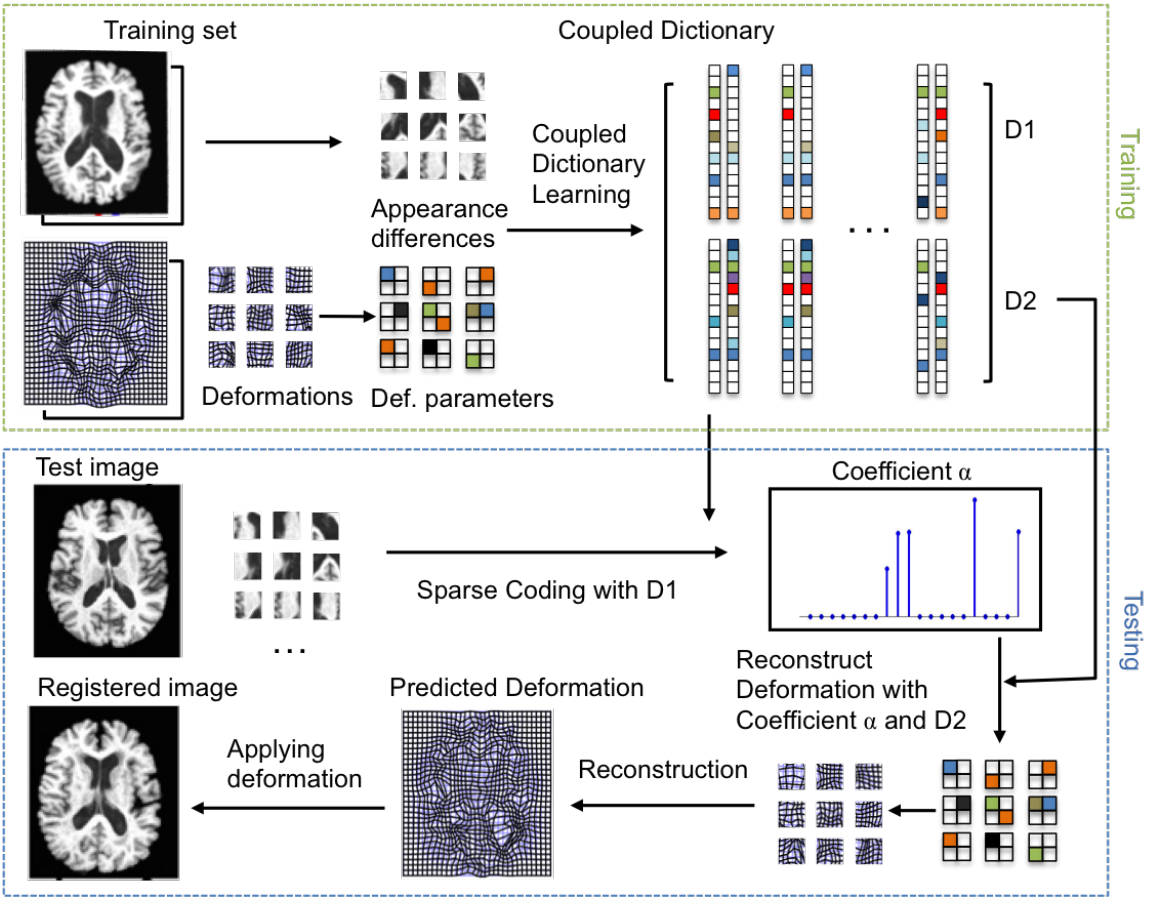


Figure 4.1: Framework of proposed method. In the training phase, I learn the coupled dictionary from training difference images and their corresponding deformations. In the testing phase, I obtain the coefficients for sparse coding of the difference image, and then predict the deformation using the coefficients and the dictionary corresponding to the deformation. Finally applying the deformation to an atlas image results in a registered atlas to a test image.

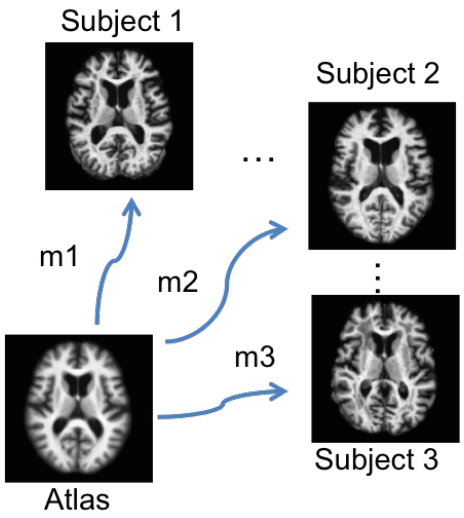


Figure 4.2: Illustration of training set and atlas. Here, $m_i,\ i = 1, \ldots, n$ are deformations generated by atlas construction (Singh et al., 2013).

where I_i is a patch of I. Equation (4.1) is a 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. I summarize my algorithms in Algorithm 3 and Algorithm 4.

Algorithm 3 Semi-Coupled Dictionary Learning

Input: t_i^1 and t_i^2, the corresponding patches from difference images and deformations. Initial $D^1 = \{d_1^1, \ldots, d_m^1\}$ and $D^2 = \{d_1^2, \ldots, d_m^2\}$ with m random t_i^1 and t_i^2 pairs, and initial W with identity matrix.

Output: D^1, D^2, W

while not converged **do**

1. Fix D^1, D^2 and W, update α_i^1 and α_i^2 in Eq. (2.13);
2. Fix α_i^1, α_i^2 and W, update D^1 and D^2 in Eq. (2.14);
3. Fix α_i^1, α_i^2, D^1 and D^2, update W in Eq. (2.16);

end while

Algorithm 4 Deformation Estimation

Input: D^1, D^2, W, I_i patches from a difference image.
Output: ϕ
for each i **do**
1. Obtain α_i^1 by solving Eq. (4.1);
2. Estimate $p_i = D^2 W \alpha_i^1$;
3. Obtain ϕ_i from p_i;

end for
Estimate ϕ based on all the ϕ_i.

4.3 Deformation parametrization

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.[1] Three commonly used deformation parametrization methods are introduced in the following sections.

4.3.1 Displacement

For any deformation ϕ, we can represent it as an identity map and a displacement at each pixel (or voxel) position,

$$\phi(x) = \mathtt{id}(x) + U(x),$$

where $\mathtt{id}$ is the identity map, i.e. $\mathtt{id}(x) = x$. The displacement field can be estimated by registration methods such as optical flow and the demons algorithm (Horn and Schunck, 1981; Pennec et al., 1999). In order to reduce the dimensionality, principal component analysis (PCA) can be applied to the displacement U (Chou et al., 2013; Kim et al., 2012), and then the deformations are parametrized as the coefficients of the principal components. PCA is beneficial for generating compact representations for deformations which show certain patterns such as respiratory lung motion (Chou et al., 2013).

[1] Assume there is a perfect mapping.

4.3.2 B-spline transformation

Without loss of generality, let $x = \{x_1, x_2, x_3\}$ be a 3d vector to represent a location in a 3d image volume. Let C denote a $n_{x_1} \times n_{x_2} \times n_{x_3}$ mesh of control points $c_{i,j,k}$ with uniform spacing σ. Then the B-spline transformation can be parametrized as,

$$\begin{aligned} \phi(x) &= \phi_{\text{global}}(x) + \phi_{\text{local}}(x), \\ &= \phi_{\text{global}}(x) + \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(u)B_m(v)B_n(w)c_{i+l,j+m,k+n}, \end{aligned} \tag{4.2}$$

where $i = \lfloor x_1/n_{x_1} \rfloor - 1$, $j = \lfloor x_2/n_{x_2} \rfloor - 1$, $k = \lfloor x_3/n_{x_3} \rfloor - 1$, $u = x_1/n_{x_1} - \lfloor x_1/n_{x_1} \rfloor$, $v = x_2/n_{x_2} - \lfloor x_2/n_{x_2} \rfloor$, $w = x_3/n_{x_3} - \lfloor x_3/n_{x_3} \rfloor$ and where B_k represents the kth basis function of the B-spline (Rueckert et al., 1999),

$$\begin{aligned} B_0(u) &= (1-u)^3/6 \\ B_1(u) &= (3u^3 - 6u^2 + 4)/6 \\ B_2(u) &= (-3u^3 + 3u^2 + 3u + 1)/6 \\ B_3(u) &= u^3/6. \end{aligned} \tag{4.3}$$

ϕ_{global} describes the global deformation which is usually captured by rigid or affine transformation models. If no global deformation is modeled, then $\phi_{\text{global}} = \texttt{id}$. The B-spline transformation model can represent complex deformations with a limited number of parameters for each control point.

4.3.3 Initial momentum

In the large deformation diffeomorphic metric mapping (LDDMM) framework (Miller et al., 2006), initial momentum can be used as a parametrization for the deformation. Given a source

image I_0 and a target image I_1, the LDDMM algorithm solves

$$\begin{aligned} &\min_{v_t} \frac{1}{2} \int_0^1 \|v_t\|_V^2 \mathrm{d}t + \frac{1}{\sigma^2} \|I_0 \circ \phi_{1,0} - I_1\|_2^2, \\ &\text{s.t. } \dot{\phi}_{t,0} + (D\phi_{t,0}) v_t = 0, \phi_{0,0} = \mathtt{id}, \end{aligned} \tag{4.4}$$

where v_t is the velocity vector field at time $t,\ t \in [0, 1]$. At a given time, t, v_t is a smooth vector field defined over the underlying image grid, Ω. The norm of this vector field image is defined using a differential operator, L, which also controls the smoothness of the velocity fields. In particular, $\|v_t\|_V^2 = \langle L^\dagger L v_t, v_t \rangle_2 = \int_\Omega (L^\dagger L v_t(x))^T v_t(x) dx$. The $\phi_{s,t}$ defines a map for a pixel/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_0, is dual to the initial velocity, v_0, and is defined by the linear mapping, $m_0 = (L^\dagger L) v_0$. The initial momentum completely parameterizes the geodesic connecting the source and target images. For a more in-depth discussion about LDDMM image registration and momentum based parametrization, refer to (Younes, 2010; Younes et al., 2009).

4.4 Results

I present three experiments to validate the proposed method: a synthetic experiment capturing translations only to illustrate the overall concept (Section 4.4.1), a synthetic experiment illustrating the behavior for both b-spline and diffeomorphic transformations (Section 4.4.2), and an experiment on real brain data (Section 4.4.3).

4.4.1 Experiment on synthetic data with translations

I first test the proposed method on synthetic data with simple random translations. Visual inspection of experiment's results provide insight into my method's behavior. I created a binary cross image as atlas, and then I translated the atlas with random translation vectors $(x_i,\ y_i)$[2] where $-20 \leq x_i \leq 20, -20 \leq y_i \leq 20, i = 1, \ldots, 5000$ to generate the training space (shown in Figure 4.3). This ensures that the cross is always fully contained within each training image. The image size is 64×64.

[2] (x_i, y_i) are real valued translations, for example translation by (1.5, 2.3).

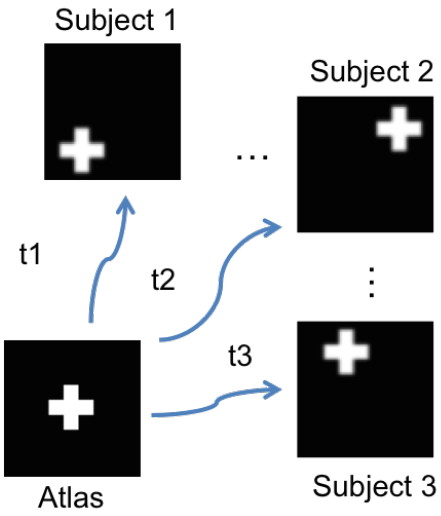


Figure 4.3: Illustration of training set generation for experiment in Section 4.4.1. Here, $t_i,\ i = 1, \ldots, n$ are translation vectors which translate the atlas.

Figure 4.4 shows an example of the training image and the atlas image. In order to capture all translations, I considered each training image as a *single* patch. The coupled dictionary was trained on difference images between the translated cross images and the atlas image as well as the corresponding translation vectors. I fixed $\lambda = 0.1$ in Equation (2.11)) during the experiment. I used translation errors (Euclidean distances between original and predicted translation vectors) and Dice coefficient[3] to evaluate my method. Let A, B be two sets of points, the Dice coefficient is defined as, $s = (2|A \cap B|)/(|A| + |B|)$. If $A = B$, the Dice coefficient $s = 1$, similarly, if $A \cap B = \varnothing$, the Dice coefficient $s = 0$.

I tested my method on 100 test images with random translations (generated separately from the training images). Figure 4.5 illustrates that a dictionary size larger than 100 obtains Dice coefficients larger than 90% and translation errors smaller than 1 (subpixel accuracy), demonstrating good deformation prediction results.

Figure 4.6 and Figure 4.7 illustrate two examples of how the dictionary atoms are used to reconstruct the test images for both SCDL and CDL methods. The dictionary atoms are more and

[3]Due to the fact that the atlas and translated images in this test are binary images, I can use Dice coefficient to evaluate the performance of my method.

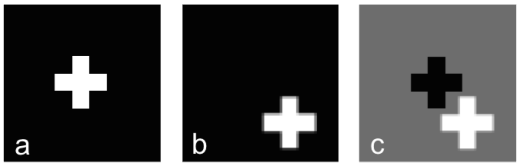


Figure 4.4: Translation experiment: Illustration of synthetic (a) atlas image, (b) translated training image, (c) difference image between atlas and translated images. (Intensities in (c) scaled for visualization.)

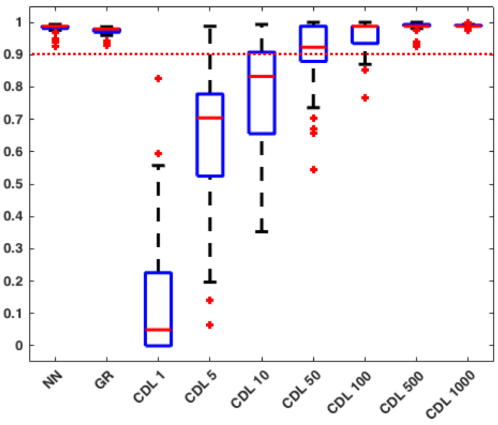
more blurry when decreasing the number of dictionary atoms which also results in less precise image reconstruction results. However, although the quality of the reconstruction of the image differences is decreasing for less than 500 dictionary atoms, the deformation prediction can still achieve good results (even when the dictionary has only 100 or 50 dictionary atoms). Both dictionary learning methods demonstrate their capability to reduce the search space (from 5000 to about 100) at the same time retaining the prediction accuracy in this experiment. In this test, nearest neighbor search (NN) (reconstructing the image patch with its nearest neighbor in the training image) and global regression (Chou et al., 2013) can also achieve good results, because the sampling was dense (5000 samples for about 40×40 possible integer translation vectors [4]) of the translation space. The results of SCDL and CDL are almost the same for this experiment because the transformation model was simple (just translations) here. I will show that SCDL outperforms CDL in the following experiments.

4.4.2 Experiment on synthetic data with local deformations

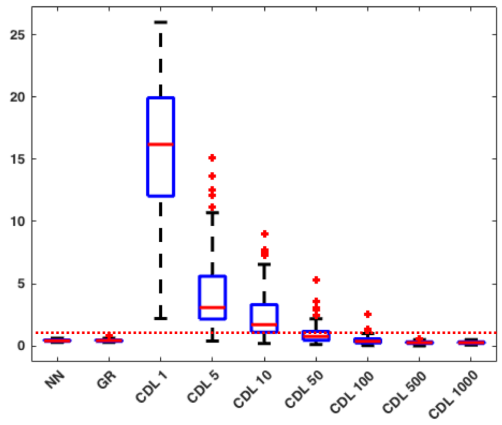
4.4.2.1 Local deformation with random b-spline parametrization

I tested the proposed method on more general deformations with B-spline transformations. This transformation model is more complex than pure translation. I created a smoothed cross image (in Figure 4.9) as atlas. The synthetic subject images are generated by applying B-spline transformations (Rueckert et al., 1999). I generated coefficients for 3×3 control points by randomly sampling from a Gaussian distribution with $\sigma = 0.5$. Thus the parametrization for the deformation is

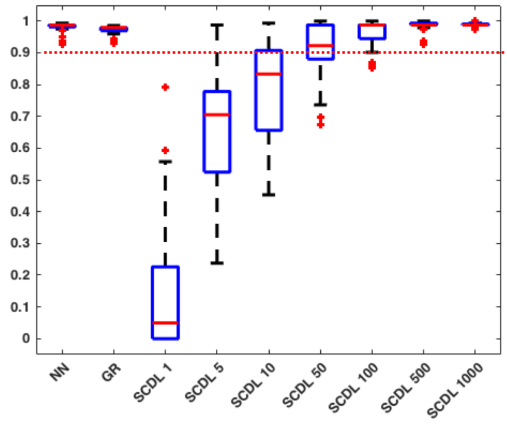
[4] I sampled the translation space with real numbers thus the number of real translation vectors is larger than the number of integer translation vectors.



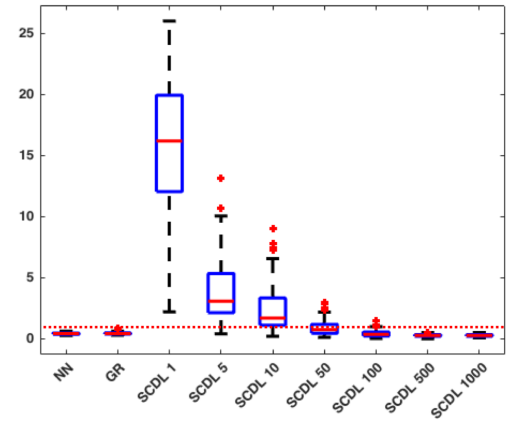
(a) Dice coefficients for CDL



(b) Translation errors for CDL



(c) Dice coefficients for SCDL



(d) Translation errors for SCDL

Figure 4.5: Deformation prediction results for expriment on synthetic data with random translations in Section 4.4.1. (a,c) show the boxplot for the Dice coefficients (the red dashed line indicates Dice coefficient = 0.9) and (b,d) show the boxplot for mean translation errors in pixels (the red dashed line indicates translations = 1 pixel) for all the test images with respect to different dictionary size for CDL and SCDL methods respectively. NN indicates nearest neighbor search method. GR indicates global regression method (Chou et al., 2013).

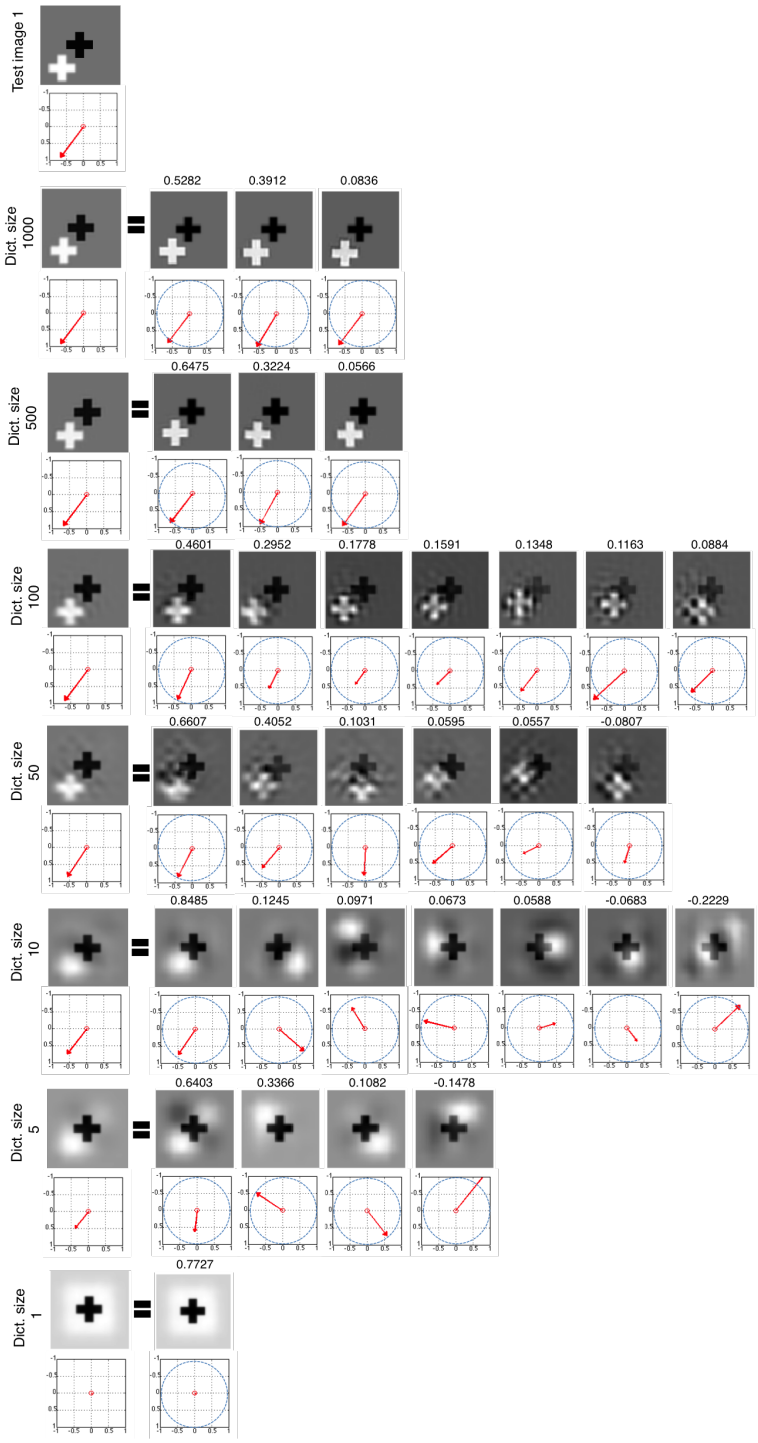


Figure 4.6: Image reconstruction results for experiment on synthetic data with random translations with respect to different dictionary size for test image using CDL for experiment in Section 4.4.1. The number on each dictionary item indicates the coefficient for the item to reconstruct the test image. The red vectors under the cross images are the corresponding translation vectors. The blue dashed circle indicates the unit circle. Note that the translation vectors for test images are unknown in testing phase, thus my method predicts the translation vectors only based on appearances. Here the translation vectors are not on the unit circle because they are jointly normalized with image residuals during dictionary learning.

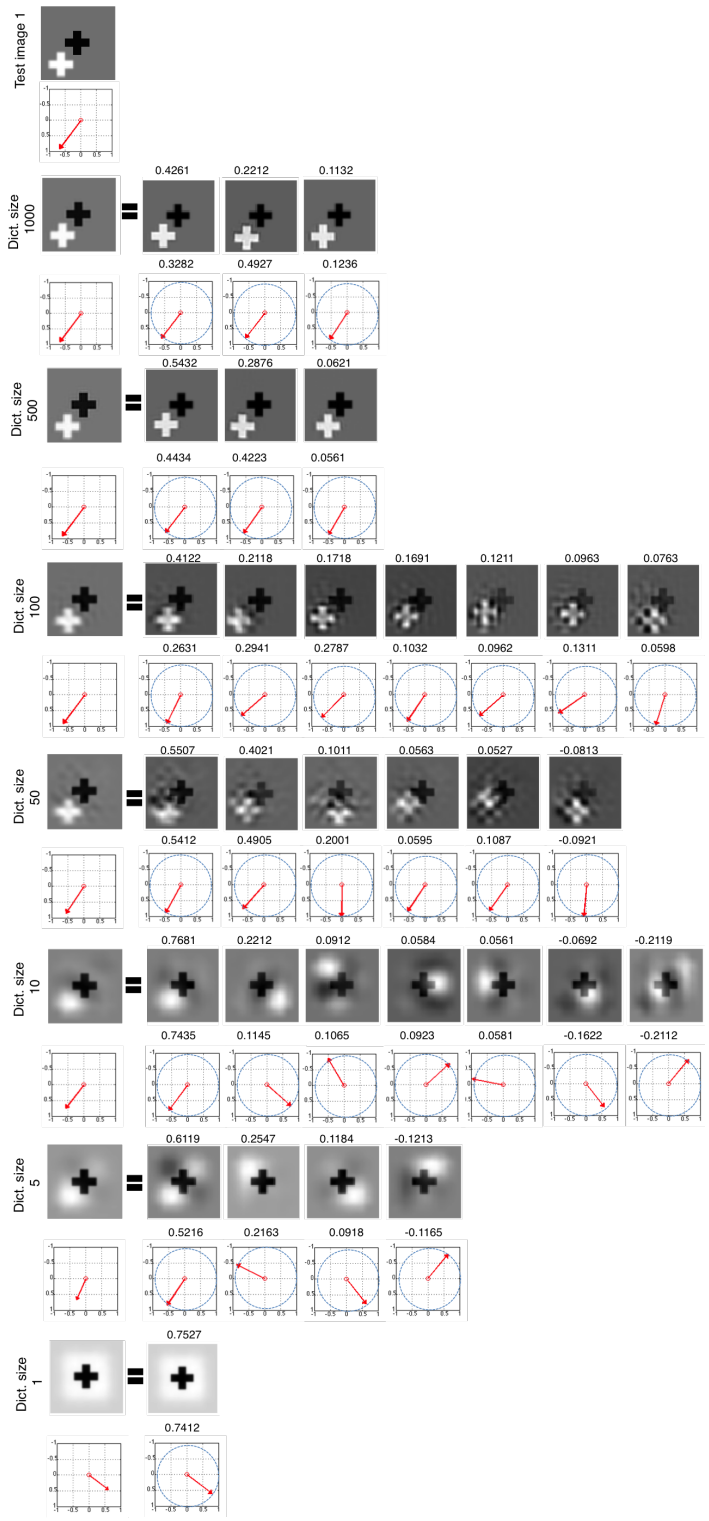


Figure 4.7: Image reconstruction results with different dictionary size for test image using SCDL for experiment in Section 4.4.1. The number on each dictionary item indicates the coefficient for the item to reconstruct the test image. The red vectors under the cross images are the corresponding translation vectors. The blue dashed circle indicates the unit circle. Note that the translation vectors for test images are unknown in testing phase, thus my method predicts the translation vectors only based on appearances.

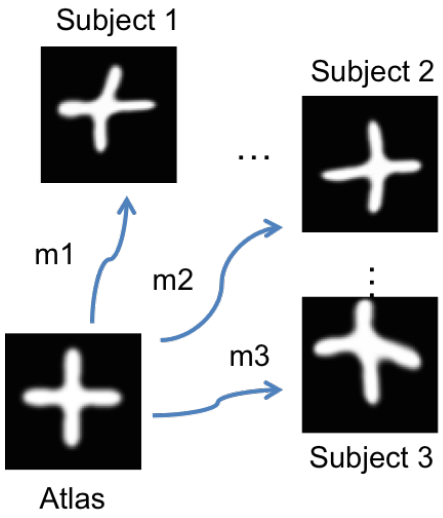


Figure 4.8: Illustration of training set generation by shooting with initial momentum (Singh et al., 2013) for the experiment in Section 4.2.2. Here, $m_i,\ i = 1, \ldots, n$ are initial momenta obtained from LDDMM by registering atlas to subject images in Section 4.2.1. The subject images are generated by shooting the atlas image with initial momenta.

the parameters of the control points for the B-spline transformation. I trained the coupled dictionary on difference images and B-spline transformation parameters. The dictionaries are learned on the whole training images. I used 1000 training images with image size 128×128, and learn the dictionaries with different numbers of atoms. I compared my introduced SCDL method with CDL, NN and global regression (GR) (Chou et al., 2013) methods on deformation prediction for 100 test images. The results are shown in Figure 4.10 and Table 4.1. Both SCDL and CDL methods demonstrate better performance compared with the NN and GR methods. Given a dictionary size of 1000, the whole training set can be exactly represented by the dictionary (1000 images, and each image for one column). However, the results for dictionary size with 1000 for both methods show that the performance does not improve significantly (p-value$_{SCDL} = 0.96$[5], p-value$_{CDL} = 0.98$) compared with the dictionary size equal to 500. This demonstrates the power of my method to reduce the search space while at the same time maintaining the accuracy. Note that SCDL further improves the performance over CDL method however the improvements are limited.

[5]The p-value is based on the paired t-test, and the null hypothesis is that the means of the paired samples are equal.

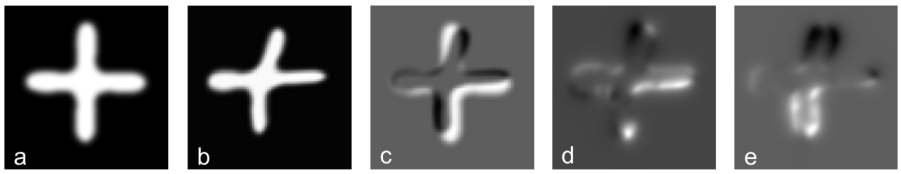
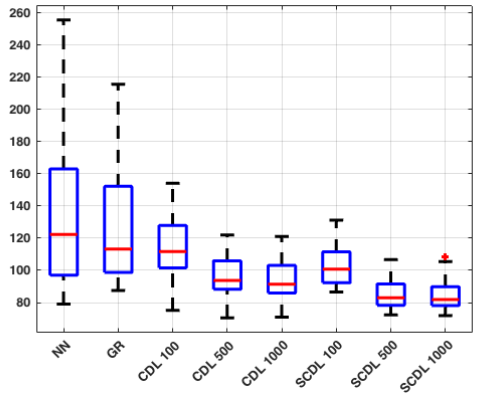
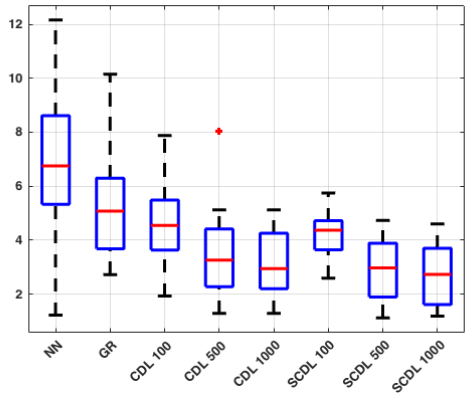


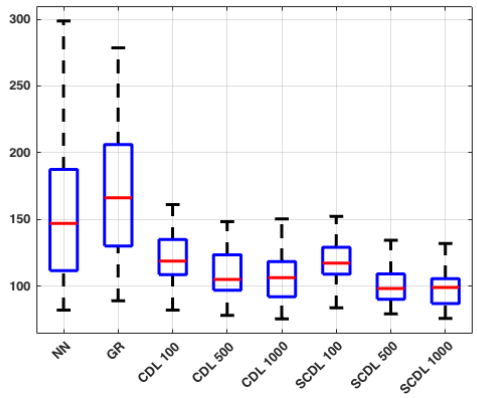
Figure 4.9: Local deformation experiment (B-Spline): Illustration of synthetic (a) atlas image, (b) deformed training image, (c) difference image between atlas and deformed images and (d, e) corresponding initial momentum in x and y direction respectively for experiment in Section 4.2.2. (Note that the intensities in (c), (d) and (e) are scaled for better visualization.)



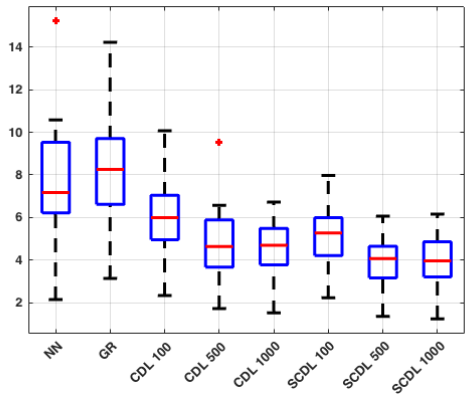
(a) SSD of images



(b) MAE of deformations



(c) SSD of images



(d) MAE of deformations

Figure 4.10: Results of experiments for B-spline transformation and initial momenta parametrization in Section 4.3.1 and 4.3.2 respectively. (a,b) shows the registration results by predicting B-spline transformation parameters experiment in Section 4.3.1 while (c,d) illustrate the results of registration by predicting initial momenta experiment in Section 4.3.2. (a,c) shows the boxplot of sum of squared differences (SSD) between deformed test images and atlas images with predicted deformation, (b,d) shows the mean absolute errors (MAE) (in pixels) of each pixel on the deformations for different methods. CDL means standard coupled dictionary learning method, SCDL denotes semi-coupled dictionary learning method, and the number besides the dictionary learning methods indicates the dictionary size(number of items). Image size is 128×128.

Table 4.1: Statistics of registration results by predicting B-spline parameters for B-spline transformation experiment in Section 4.3.1. 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 denotes global regression method and CDL and SCDL represent coupled and semi-coupled dictionary learning methods respectively.

Method	median	mean	min	max	std
RAW	17.9776	19.2912	6.4132	45.1841	9.1304
NN	6.7063	6.5239	1.2399	12.2234	3.0948
GR	5.0712	5.2530	2.7510	10.1439	1.9895
CDL100	4.5859	4.5988	1.9389	7.0713	1.2891
CDL500	3.1325	3.1531	1.0824	8.5439	1.3158
CDL1000	3.1904	3.1464	1.0815	5.1018	1.1676
SCDL100	4.3747	4.1061	1.2246	5.7707	0.9841
SCDL500	2.7679	2.7454	0.7532	4.7582	1.1815
SCDL1000	**2.7261**	**2.7189**	**0.7407**	**4.5759**	**1.1001**

Table 4.2: Statistics of registration results by predicting of initial momenta for initial momenta parametrization experiment in Section 4.3.2. The numbers show the mean absolute errors (MAE) in pixels of predicted deformation to the ground truth.

Method	median	mean	min	max	std
RAW	18.9566	19.8012	6.9132	45.1841	14.549
NN	7.1677	7.6787	2.1351	15.3961	2.1215
GR	8.2587	8.3353	3.1331	14.2234	2.2797
CDL100	5.8651	6.8965	2.3514	10.7019	1.6552
CDL500	4.6312	4.7447	1.7329	9.3345	1.5292
CDL1000	4.6864	4.6347	1.6185	6.1249	1.3584
SCDL100	5.2734	5.1659	2.2229	7.818	1.2951
SCDL500	4.0176	3.9974	1.3294	6.0656	1.1914
SCDL1000	**3.9767**	**3.945**	**1.3439**	**6.0132**	**1.1752**

4.4.2.2 Random transformations with initial momentum parametrization

I used a similar setting to the previous experiment. Instead of directly using training subject images generated by applying B-spline transformations, I registered the atlas image to the subject images with LDDMM registration to obtain the initial momenta and also shoot the atlas with obtained initial momenta to generate a set of new subject images as ground truth (shown in Figure 4.8). Thus I obtained a different deformation parametrization compared with B-spline transformation, i.e. initial momenta for LDDMM. The initial momenta parametrize the geodesics between source and target images, and thus the linear combination of the initial momenta (in the SDL and SCDL models) in the tangent space of the geodesic makes sense. I trained the coupled dictionary on difference images and initial momenta. The dictionaries are learned on the whole training images. I had 1000 training images with image size 128×128, and learn the dictionaries with different numbers of items. I compared my introduced SCDL method with CDL, NN and GR methods on deformation prediction for 100 test images. The results are shown in Figure 4.10 and Table 4.2. The results are similar to the results of the previous experiment for B-spline transformations. Both SCDL and CDL methods demonstrate better performance at the same time reducing the search space compared with the NN method. However, the results for dictionary size with 1000 for both methods show that the performance does not improve significantly (p-value$_{SCDL} = 0.98$, p-value$_{CDL} = 0.97$) compared with the dictionary size equal to 500 which also demonstrates that our method was able to achieve good deformation prediction with a moderately sized dictionary.

4.4.2.3 Local deformation with random initial momenta parametrization

In this experiment, I tested my methods on random local deformations. Similar to the B-spline transformation experiment, I used the smoothed cross image (in Figure 4.12) as atlas. The training data are generated by applying random deformations to the atlas (shown in Fig 4.11). The deformations are parametrized by scalar initial momentum (Singh et al., 2013). The random initial momenta are generated by random sampling from a Gaussian distribution with $\sigma = 3$ (corresponds to maximum displacement about 10 pixels).

I trained a coupled dictionary based on the intensity differences between atlas and deformed training images and the deformation's parameters, initial momenta. The image size is 128×128, and I extracted 15×15 image patches from 200 difference images to train the coupled dictionary. The image patches are extracted with 1 pixel stride, thus I obtained overlapping image patches. During the testing, I first reconstruct each local patch, then reconstruct the whole image by averaging the overlapped patches (Cao et al., 2013b). I compared the SCDL method with the CDL method, the NN method and the GR method on deformation prediction for 50 test images. Figure 4.13 and Table 4.3 show the results of deformation prediction by comparing the transformed test images with the deformed atlas image, where the parameters of the deformation were predicted by the model. Both SCDL and CDL methods show better performance compared with the NN method. However, the SCDL method does not increase the performance too much. A dictionary of size 1000 is sufficient for accurate deformation prediction, since dictionaries of size 5000 do not yield statistically significant improvement (p-value$_{SCDL} = 0.97$, p-value$_{CDL} = 0.99$). The difference between this experiment and the experiments in Section 4.4.2.1 is that I learn the dictionary based on image patches instead of whole images. In practice, it is unrealistic to have dictionary atoms of the same size as the whole image. Also, for the dictionary learning in whole image cases, the dictionaries are not over-complete, however, for image patches, the dictionaries are likely over-complete.

4.4.3 Experiment on real data

The previous experiments were based on synthetic data. Here, I tested my method on the Open Access Series of Imaging Studies (OASIS) Cross-sectional MRI data (Marcus et al., 2007). I constructed an atlas from 100 subject images with LDDMM to obtain the initial momenta (shown in Figure 4.14) (Singh et al., 2013). Figure 4.15 shows the atlas image and an example of the subject image. The coupled dictionary is learned based on the intensity difference between atlas and subject images and the corresponding initial momenta. The image size is 128×128, and I extracted 15×15 image patches from training images. I tested my method on 50 test subject images. Figure 4.16 and Table 4.4 show the results of deformation prediction by comparing the deformed

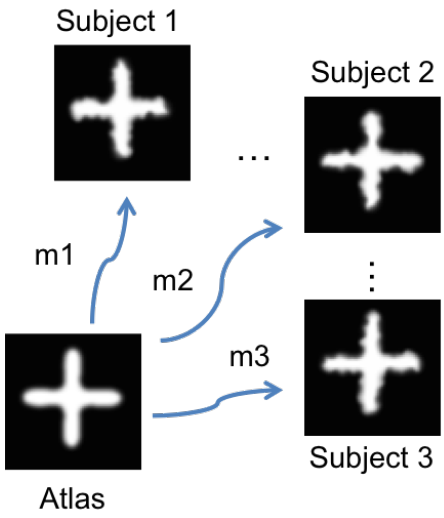


Figure 4.11: Illustration of training set generation by shooting with random initial momentum for the experiment in Section 4.3.3. Here, $m_i,\ i = 1, \ldots, n$ are randomly generated initial momenta.

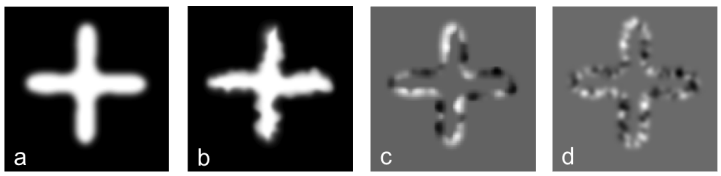
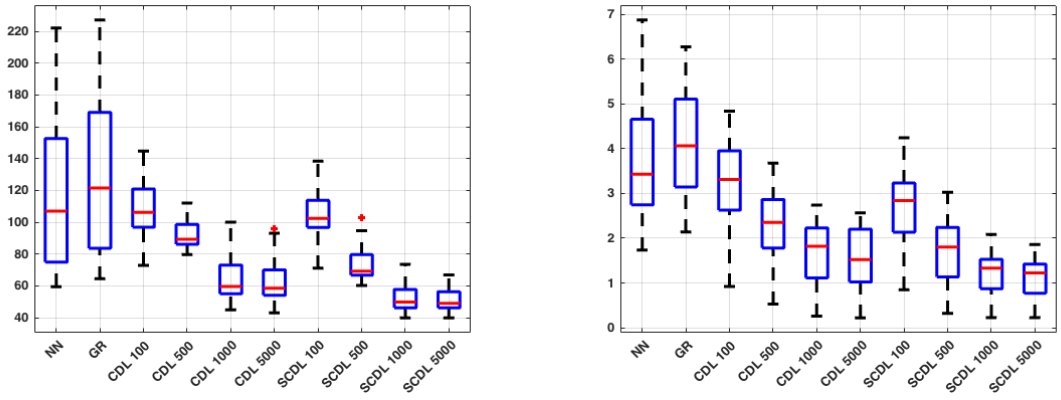


Figure 4.12: Local deformation experiment: Illustration of synthetic (a) atlas image, (b) deformed training image, (c) difference image between atlas and deformed images and (d) corresponding initial scalar momentum for experiment in Section 4.3.3. (Note that the intensities in (c) and (d) are scaled for better visualization.)



(a) SSD of images

(b) MAE of deformations

Figure 4.13: Results of local deformation with random initial momenta parametrization experiment in Section 4.3.3. (a) shows the boxplot of sum of squared differences (SSD) between deformed test images and atlas images with predicted deformation (initial momentum), (b) shows the mean absolute errors (MAE) (in pixels) of each pixel on the deformations for different methods. CDL means standard coupled dictionary learning method, SCDL denotes proposed semi-coupled dictionary learning method, and the number besides the dictionary learning methods indicates the dictionary size (number of atoms).

Table 4.3: Statistics of registration results by predicting random initial momenta for synthetic data (experiment in Section 4.3.3). 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 denotes global regression method and CDL and SCDL represent coupled and semi-coupled dictionary learning methods respectively.

Method	median	mean	min	max	std
RAW	6.0637	6.1669	3.7536	9.5566	1.7194
NN	3.4307	3.6619	1.7388	6.8764	1.9108
GR	4.0644	4.1099	2.1364	6.2764	1.5735
CDL100	3.3148	3.2546	0.9239	4.8432	1.3248
CDL500	2.3544	2.3155	0.5316	3.681	1.0838
CDL1000	1.8192	1.6761	0.262	2.7386	1.1135
CDL5000	1.5251	1.5703	0.222	2.5676	1.1808
SCDL100	2.841	2.6458	0.8508	4.2449	1.0992
SCDL500	1.8053	1.7281	0.3199	3.0272	1.111
SCDL1000	1.3331	1.2328	0.2287	2.0823	0.7631
SCDL5000	**1.2253**	**1.1202**	**0.2231**	**1.8582**	**0.6538**

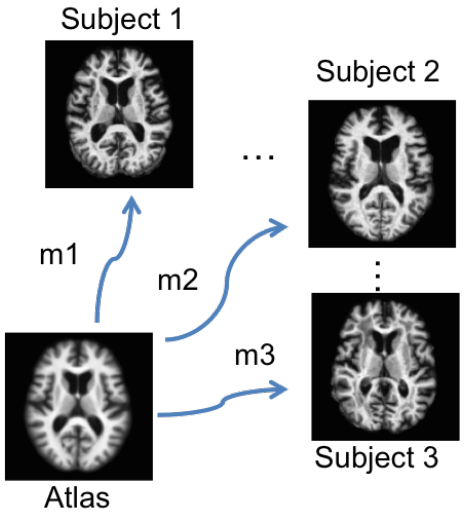


Figure 4.14: Illustration of training set and atlas for experiment on OASIA dataset in Section 4.4.3. Here, $m_i,\ i = 1, \ldots, n$ are initial momenta generated by atlas construction (Singh et al., 2013).

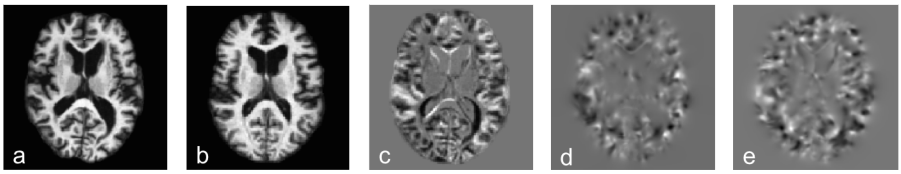
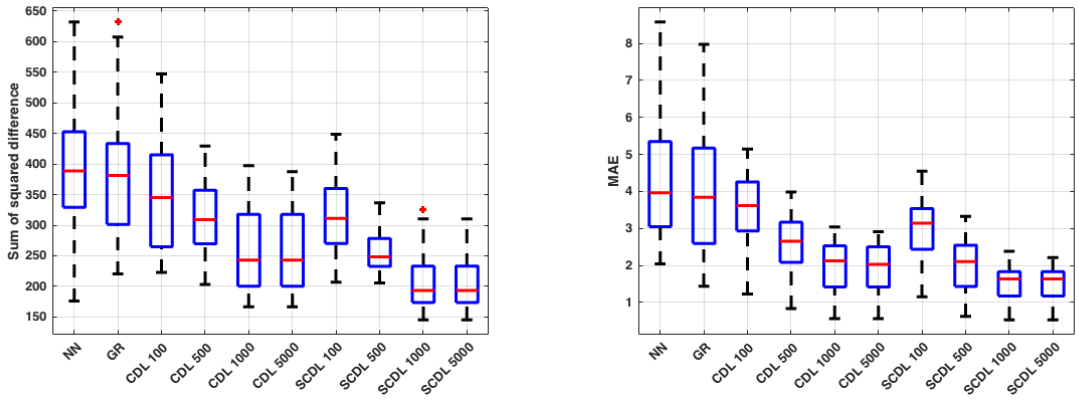


Figure 4.15: Illustration of brain (a) atlas image, (b) subject image, (c) difference image between atlas and subject images and (d,e) corresponding initial momentum in x and y direction respectively for experiment on OASIA dataset in Section 4.4.3. (Note that the intensities in (c), (d) and (e) are scaled for better visualization.)

test images with the predicted deformation to the atlas images. The SCDL method shows better performance compared with NN, GR and CDL methods, while at the same time reducing the search space compared to the NN method. Also, the dictionary size 1000 is sufficient for deformation prediction since increasing dictionary size to 5000 does not significantly (p-value$_{SCDL} = 0.98$, p-value$_{CDL} = 0.98$) improve the performance.

4.5 Discussion and conclusion

I 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 par-



(a) SSD of images

(b) MAE of deformations

Figure 4.16: Results of experiment on OASIA dataset in Section 4.4.3. (a) shows the boxplot of sum of squared differences (SSD) between deformed test images and atlas images with predicted deformation (initial momentum), (b) shows the mean absolute errors (MAE) (in pixels) of each pixel on the deformations for different methods. CDL means standard coupled dictionary learning method, SCDL denotes proposed semi-coupled dictionary learning method, and the number besides the dictionary learning methods indicates the dictionary size(number of items). Image size is 128×128.

Table 4.4: Statistics of registration results by predicting initial momenta for OASIS dataset (experiment in Section 4.4.3. The numbers show the mean absolute errors (MAE) in pixels of predicted deformation to the ground truth.

Method	median	mean	min	max	std
RAW	7.5166	7.4951	5.2495	9.9351	1.1859
NN	3.9619	4.3453	2.0205	8.5739	1.7895
GR	3.8719	4.0353	1.4205	7.9439	1.7254
CDL100	3.6301	3.5546	1.2149	5.1219	0.9597
CDL500	2.6255	2.6456	0.8432	3.981	0.7676
CDL1000	2.1182	1.9674	0.6102	3.0101	0.6896
CDL5000	2.0996	1.9308	0.6022	2.9101	0.6517
SCDL100	3.1454	2.9573	1.1829	4.5319	0.8165
SCDL500	2.1048	2.0209	0.6189	3.3272	0.7513
SCDL1000	1.6386	1.5384	0.5851	2.3821	0.6286
SCDL5000	**1.6012**	**1.5154**	**0.5841**	**2.3289**	**0.6168**

ticular, 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. GR performs comparable to the NN method and worse than CDL and SCDL. A possible reason is that the relationship between the difference image and the deformation is non-linear for complex transformation models (b-spline and LDDMM). 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. This independent normalization is important for the coupled dictionary learning of the data from two spaces with large differences, for example the difference image and deformation parameters. Joint normalization treats the data from two different spaces equally based on the assumption that the data from two spaces have the same variance which may not be the case for the data in my experiments.

Another benefit of the dictionary learning based methods are the compression of search space compared with NN method. NN method performs as well as or even better compared to the proposed dictionary learning methods for the translation experiment. However, its performance is worse in other deformable transformation experiments. The main reason is that in the translation case, I use a dense sampling of the translation space while in the other experiments it is prohibitive to obtain such a dense sampling of the deformation space which is usually high dimensional.

The proposed SCDL method can be used for modeling data in different spaces with large discrepancies. It provides a more flexible modeling of the coupled data compared with CDL.

CHAPTER 5: COUPLED DICTIONARY LEARNING FOR GTPASE ACTIVITIES AND CELL MOVEMENTS

Exploring the spatio-temporal coordination of GTPase activities and cell movements is crucial for understanding cell dynamics, for example, the initiation and termination of protein biosynthesis (Scheffzek and Ahmadian, 2005). In this chapter, the spatio-temporal relationship between GTPase activations and cell movements is investigated. Previous research in (Jaffe and Hall, 2005; Ridley et al., 2003; Burridge and Wennerberg, 2004) has shown that GTPases are activated at the edge of moving cells. Thus in this chapter, only the GTPase activations and cell movements on the cell boundary are studied. The first question I try to answer is: Do the GTPase activations relate to cell movement? To answer this question, I first apply enrichment analysis to test whether particular patterns of activations tend to co-occur along with patterns of cell movement. If some groups of cell movements show enrichment in some groups of activations, it indicates that for these groups, the movements are related to activations, and therefore suggests the possibility of predicting cell movements from activations. Hence, I investigate whether movements can be predicted from activations. A coupled dictionary learning method is applied to learn a coupled basis for the prediction of cell movements from GTPase activations.

5.1 Introduction

GTPases (Rac1, RhoA and Cdc42), a family of enzymes that can bind GTP, play important roles in cell dynamics, such as controlling cytoskeleton dynamics (Machacek et al., 2009; Ridley et al., 2003). Cell dynamics are related to many biological processes such as tissue repair and regeneration, and disease progression of cancer (Ridley et al., 2003).

Previous research (Machacek et al., 2009; Ridley et al., 2003) shows that all three GTPases are activated at the edge of migrating cells, and there exists a spatio-temporal relationship between

GTPase activities and cell edge movements. Studying the relationship between GTPase activities and cell edge movements provides a way to understand the mechanism of cell dynamics thereby resulting in a better interpretation of biological processes. GTPase activities can be measured with biosensor imaging of mouse embryonic fibroblasts (MEFs) (Machacek et al., 2009). Here, the intensities of the biosensor images for MEFs indicate the activity of GTPases, i.e. the current of GTPases.

However, learning the relationship between GTPase activities and cell movements is challenging for the following reasons: (i) a good data representation is required to capture the relationship between GTPase activities and cell movements; (ii) establishing the spatio-temporal correspondences between GTPase activities and cell movements is difficult as both the shapes and appearances (activations) of cells change during cell movements.

The spatio-temporal coordination between GTPase activities and cell movements is explored in (Machacek et al., 2009). However, this approach has some drawbacks,

- the studied regions of MEFs are manually selected by experts;
- the spatial interactions between neighboring points are also ignored.

This chapter investigates the relationship between GTPase activities and cell movements. One simple relationship between GTPase activities and cell movements can be uncovered by enrichment analysis. It can determine whether groups of movements are enriched in activities. Data points can then be selected based on enrichment analysis results. To further investigate the predictability of movements from activities, a coupled dictionary learning method is used to learn common spatial and temporal patterns for GTPase activation and cell movements. Fig 5.1 shows the flowchart of the proposed method for this chapter.

This chapter is organized as follows: Section 5.2 describes the methods I use to analyze the GTPase activation and cell movement data. Section 5.3 illustrates the results of the proposed method. The chapter concludes with a summary of results and an outlook on future work in Section 5.4.

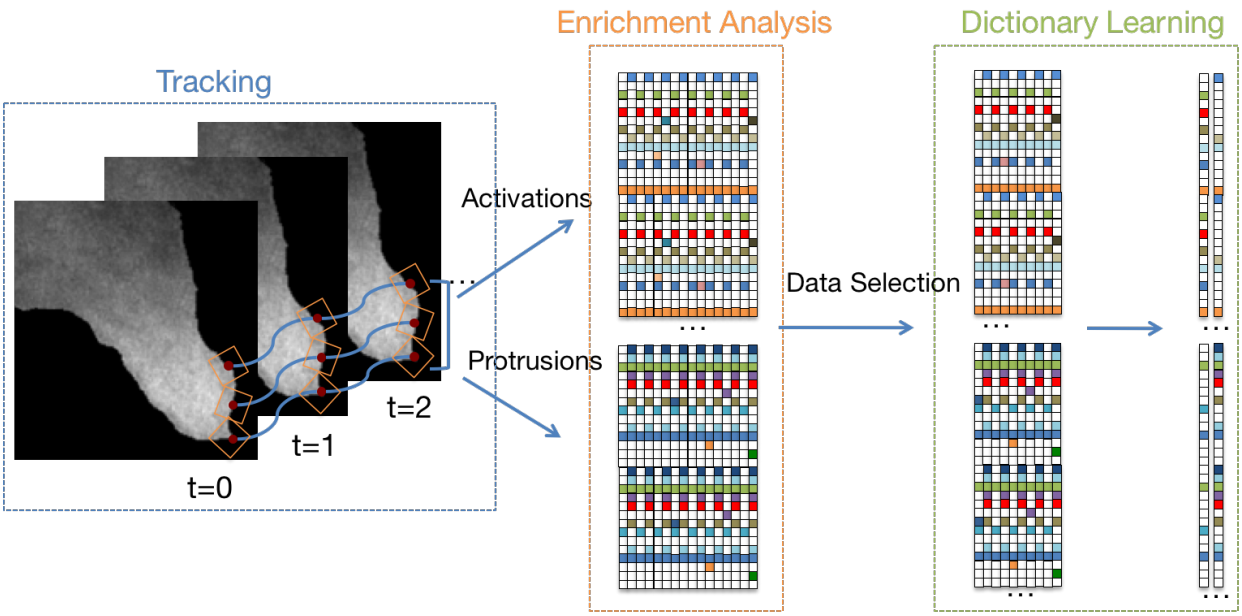


Figure 5.1: Flowchart for coupled dictionary learning for activations and protrusions.

5.2 Method

Before analyzing the relationship between GTPase activities and cell movements, the data relating the information between GTPase activation and cell movements needs to be extracted from video frames.

5.2.1 Cell Edge Movement Measure

Measuring cell edge movements can be achieved by sampling a set of markers on the cell boundary, and tracking the markers over time (Machacek and Danuser, 2006). The velocity v of a marker is defined as the signed displacement d between two consecutive frames, i.e. $v = d/t$. The cell shape at different time points can be recovered by integrating the velocities of the markers over time from the initial positions of the markers. As a result, the velocities of markers are measured as a proxy for the cell shape.

5.2.1.1 Definition of Protrusion and Retraction

Similar to the methods in (Machacek and Danuser, 2006; Machacek et al., 2009), protrusions and retractions are only considered in the direction normal to the cell boundary (Machacek and

Danuser, 2006)[1] where positive velocity corresponds to protrusion and negative velocity corresponds to retraction. Note that the definition ignores the point movements which are not perpendicular to the boundary and may not be an exact estimation of the point movements. An alternative definition could include non-perpendicular motions, however, this will make the point tracking more challenging because it is difficult to find proper features about the boundary points to track (both shapes and appearances (intensities) of the cell boundary change of time).

5.2.1.2 Boundary Tracking

Due to the large intensity difference between background and foreground of the cell in each frame, intensity thresholding is used to segment the cell. In order to obtain a smooth boundary, the segmented binary image is first smoothed with a Gaussian filter. The boundary of the cell is then extracted from the 0.5 level set. Figure 5.2 shows an example for the boundary extraction. Given the boundary Γ_t at time t, a set of markers $p_i, i \in \{1, \ldots, n\}$ are sampled with equal distance on Γ_t. The distance is set to 7 pixels where the pixel size is $330nm$. A level set function ϕ_t is defined as the signed distance function with respect to Γ_t.

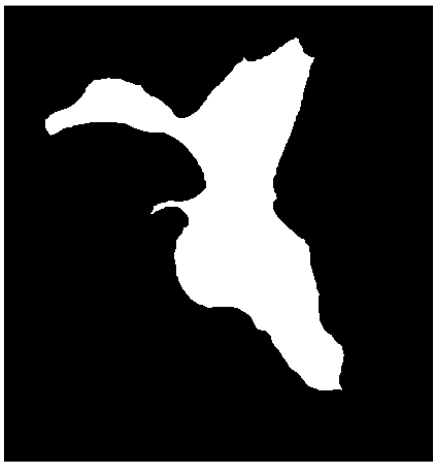
I use a simple tracking method by directly searching the corresponding markers along the normal directions of the cell boundary. Let Γ_t, Γ_{t+1} be the cell boundaries in two consecutive frames at time t and $t+1$ respectively. ϕ_t and ϕ_{t+1} are level set functions with respect to Γ_t and Γ_{t+1}. The value of ϕ_t at location (x, y) is the distance to the closest point of the boundary Γ_t (Figure 5.3). The distance values at the locations inside the cell boundary have negative values and at the locations outside have positive values.

For a marker $p_i = (x_{p_i}, y_{p_i})$ on boundary Γ_t, $\nabla\phi_t(x_p, y_p)$ is the gradient of the level set ϕ_t at point p_i, thus the direction of $\nabla\phi_t(x_p, y_p)$ points in the direction with greatest increase of the distance of ϕ_t.

The tracking process can be achieved by a line search along the normal direction for each marker on the cell boundary as illustrated in Figure 5.4. The velocity v_i^t at marker p_i is defined as the distance d between p_i and q_i on Γ_t and Γ_{t+1} respectively which is computed by the tracking process

[1]This definition of protrusion and retraction simplifies the tracking process (Machacek and Danuser, 2006).

(a) Raw image



(b) Segmentation result

(c) Smoothed segmentation result

(d) Boundary extraction result

Figure 5.2: An example of boundary extration.

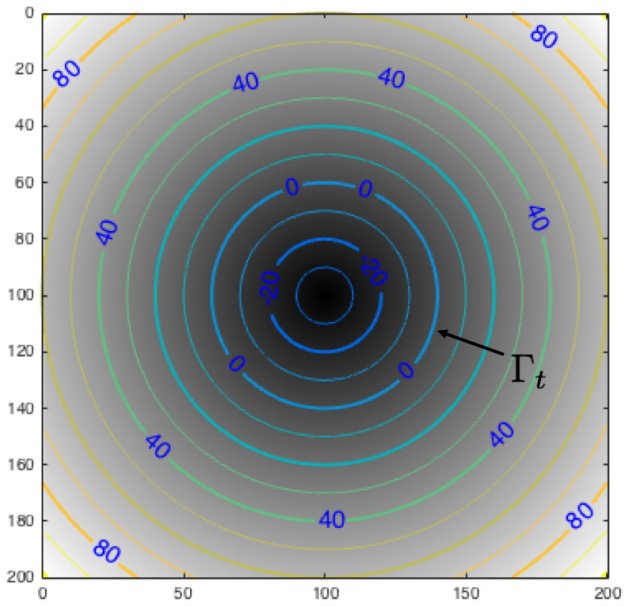


Figure 5.3: Example of the level set function ϕ_t for a boundary Γ_t. The intensity values at each pixel indicates the distance to the boundary Γ_t. The points on each circle which are superimposed on ϕ_t have the same distance to the boundary. The numbers on the circles indicate the distances to the boundary with positive values outside the boundary and negative values inside the boundary.

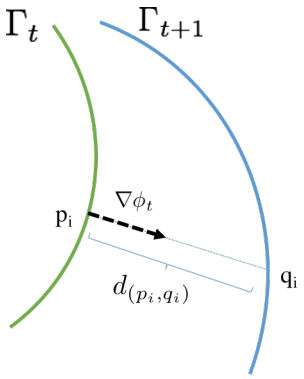


Figure 5.4: Illustration of boundary tracking. p is the location of a marker on the boundary Γ_t, and the goal is to find the corresponding location q of the marker on Γ_{t+1}. $\nabla\phi_t$ is the gradient direction of ϕ_t.

divided by the time between frame t and $t+1$ (set to one for my data for simplicity). Positive velocity indicates cell protrusion. The tracking algorithm is described in Appendix A.1. After boundary tracking, the velocities are extracted for each marker at different frames $\mathbf{v}_i = \{v_i^1, \ldots, v_i^{k-1}\}$ where k is the total number of frames.

5.2.2 GTPase Activity Measure

As the relationship between GTPase activities and cell movements is concentrated at the cell boundary (Machacek et al., 2009), GTPase activities are only extracted from the cell boundary. Similar to (Machacek et al., 2009), sampling windows on the boundary Γ_t are defined as a 3 pixel deep band (about 0.9 μm) at the cell edge with width of 7 pixels (about 2 μm) which are centered at the markers during boundary tracking. The GTPase activations (Rac1, Cdc42 and RhoA) for each sampling window can be obtained by averaging the respective activations inside each sampling window. Thus I can obtain $\mathbf{a}_{(Rac1)i} = \{a^1_{(Rac1)i}, \ldots, a^{k-1}_{(Rac1)i}\}$, $\mathbf{a}_{(Cdc42)i} = \{a^1_{(Cdc42)i}, \ldots, a^{k-1}_{(Cdc42)i}\}$ and $\mathbf{a}_{(RhoA)i} = \{a^1_{(RhoA)i}, \ldots, a^{k-1}_{(RhoA)i}\}$.

5.2.3 Enrichment Analysis between GTPase Activities and Cell Edge Movements

In (Machacek et al., 2009), the authors investigate the coordination between GTPase activities and cell edge movements by measuring the time shift between GTPase activations and cell velocities when they achieve maximum cross-correlation[2]. Both Rac1 and Cdc42 are activated after the edge movement with a delay of 40 seconds (10 seconds per frame) while RhoA is activated at the cell edge synchronous with the cell movement (Machacek et al., 2009). In this chapter, I only focus on $\mathbf{a}_{(RhoA)i}$ in the analysis, but the method can also be applied to other GTPase types. I use $\mathbf{a}_i$ to represent $\mathbf{a}_{(RhoA)i}$ in the following sections.

In this section, I first consider a simple relationship between activations and velocities. If the activations or velocities for some boundary points are similar to each other, a group can be formed for these boundary points. A priori it is not clear if the groups based on activations will have any relationship to the groups based on velocities. Enrichment analysis can determine if groups share

[2]I will revisit this relationship in Section 5.3.2.1

points with each other. If they do, it suggests that prediction of velocities from activations may be possible.

5.2.3.1 Enrichment Analysis

Enrichment analysis is a method commonly used in gene or protein analysis to test whether classes of genes or proteins are enriched in a set of genes or proteins (Subramanian et al., 2005). This method uses a statistical approach to identify significantly enriched groups of genes. Given the group information about GTPase activations and cell protrusions, enrichment analysis can identify the groups of activations which are enriched in the groups of protrusions. The benefits of enrichment analysis are as follows,

- no assumptions about the linear or nonlinear relationships between data points;
- easy to compute and interpret.

However, enrichment analysis needs group information of the data which is not available for $\mathbf{a}_i$ and $\mathbf{v}_i$. Therefore I use k-means clustering to determine groups.

5.2.3.2 K-means Clustering

K-means clustering aims to assign n data samples into k clusters so that each sample belongs to the cluster with the closest mean.This problem is NP-hard, however, there are efficient algorithms that are commonly used and converge quickly to a local optimum (Kanungo et al., 2002).

Given n data samples $X = \{x_i, x_2, \ldots, x_n\}$, where $x_i \in \mathbf{R}^d$ is a d-dimensional vector, k-means clustering divides the n data samples into k groups $S = \{s_1, \ldots, s_k\}$, where $k \leq n$, and minimizes the following within-cluster sum of squares,

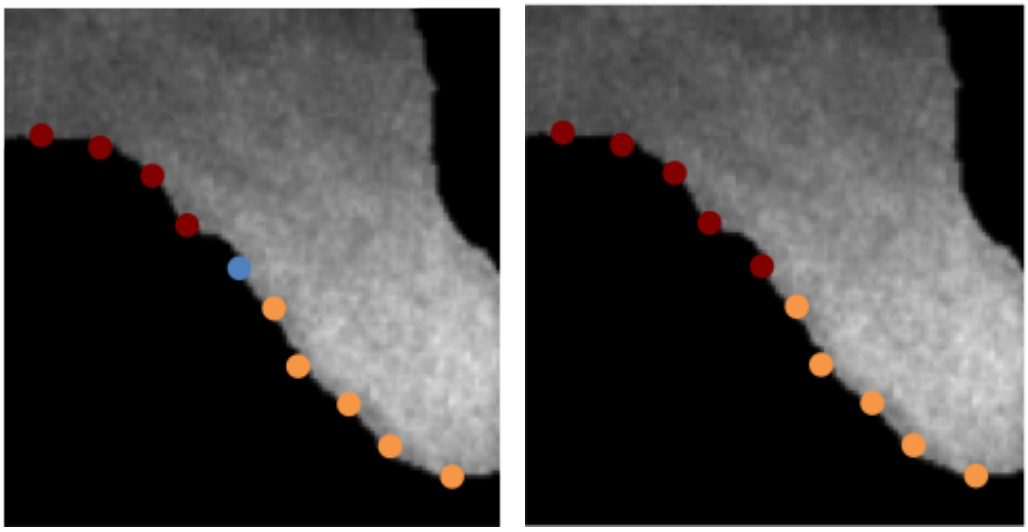
$$\text{minmize} \sum_{i=1}^{k} \sum_{j \in s_i} \|x_j - c_i\|^2, \tag{5.1}$$

where c_i is the mean of the data points in cluster s_i. The k-means clustering algorithm is described in Algorithm 5.

5.2.3.3 Data Concatenation

Applying k-means clustering on $\mathbf{a}_i$ or $\mathbf{v}_i$ for individual locations along the boundary may suffer from the problem that neighboring points get assigned to different clusters. Because the spatial consistency between different points is not taken into account if points are clustered individually. Clustering on data "patches" by concatenating the data for neighboring points can obtain more consistent results compared with clustering on single data points. For example, in Fig. 5.5 (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. 5.5 (b).

Another reason to use "patches" is that the goal of this chapter is spatio-temporal analysis. Clustering on individual points only uses temporal information. By using patches I can capture spatio-temporal aspects of the data. The patch size is set to 11 points in the experiments. I use $\mathbf{a}_{(c)i}$ and $\mathbf{v}_{(c)i}$ to represent the concatenated activations and velocities of "patches".



(a) clustering on single points (b) clustering on point "patches"

Figure 5.5: Example of clustering results based on single points and point "patches". Different colors indicate different clusters.

Enrichment analysis of activations and velocities can be achieved by hypergeometric testing. Hypergeometric testing provides p-values for the enrichment of velocities in activations.

Algorithm 5 K-means clustering algorithm

Input: Data points: $\{x_i\}$, $i \in 1, ..., N$;
Output: Cluster centers $\{c_i\}$, $i \in 1, ..., k$;
Cluster partitions $\{s_i\}$, $i \in 1, ..., k$.

1: Randomly select 'k' cluster centers;
2: Calculate the distance between each data point and the cluster centers;
3: Assign each data point to the cluster whose center is closest to the point. Let s_i be the set of points assigned to cluster i;
4: Recalculate the new cluster centers using:

$$c_i = \frac{1}{N_i} \sum_{j \in s_i} x_j,$$

where N_i is the number of points in ith cluster;
5: Iterate 2-4 steps until converge.

5.2.3.4 Hypergeometric Testing

Hypergeometric testing is based on the hypergeometric distribution. The hypergeometric distribution is a probability distribution that describes the probability of k successes in n samples without replacement, from a population of size N with exactly K successes, where each sample is either a success or a failure (Rice, 2006). The successes for the enrichment analysis on activations and velocities corresponds to the number of shared points in the cluster of activations and the cluster of velocities. In a test for over-representation of successes in the samples, the p-value is calculated as the probability of randomly drawing k or more successes from the population in n draws. For the enrichment analysis on activations and velocities, the p-value for the enrichment of velocities on activations is calculated as the probability that there are k or more than k points shared in the cluster with activations of size K and the cluster with velocities of size n for a total number of points is N.

A random variable X follows the hypergeometric distribution if its probability mass function is,

$$p(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \tag{5.2}$$

where $\binom{a}{b}$ is a binomial coefficient, $\binom{a}{b} = \frac{a!}{b!(a-b)!}$, for $0 \leq b \leq a$.

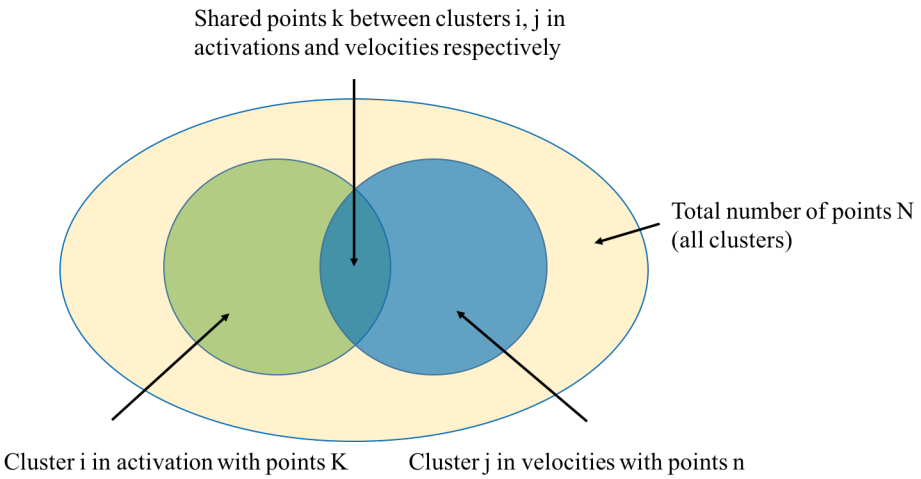


Figure 5.6: Illustration for the parameters in the hypergeometric distribution for the clusters in activations and velocities.

Assume cluster i by clustering with activations has K data samples, cluster j by clustering with velocities has n data samples, the number of points shared in both clusters i and j is k, and the total number of data samples is N. Figure 5.6 illustrate the parameters for the hypergeometric distribution for the clusters in activations and velocities.

Figure 5.7 shows the results of hypergeometric testing for clustering results based on activations (only RhoA is shown here) and velocities. The p-values are aligned to put the lowest value on the diagonal. I can see the enrichment between clusters with GTPase activation and cell velocities, for example, the enrichment of velocities in activation clusters 1, 4, 8, 9, 11, 12, 14, 15, 16 and 19 suggests that there is a relationship between activations and velocities. Figure 5.8 shows the number of shared points between clusters based on activations and velocities.

5.2.3.5 Boundary Points Selection

For the clusters that show no enrichment, it means that the the null hypothesis cannot be rejected. Thus the activations and velocities for these points may not be directly related to each other or may have some complicated relationship that may not have been detected by the enrichment test. To simplify further analysis, I only focus on the shared points in the clusters with enrichment. In these clusters, the enrichment suggests that there may be common patterns for activations and velocities and hence velocities may be predictable from the activations.

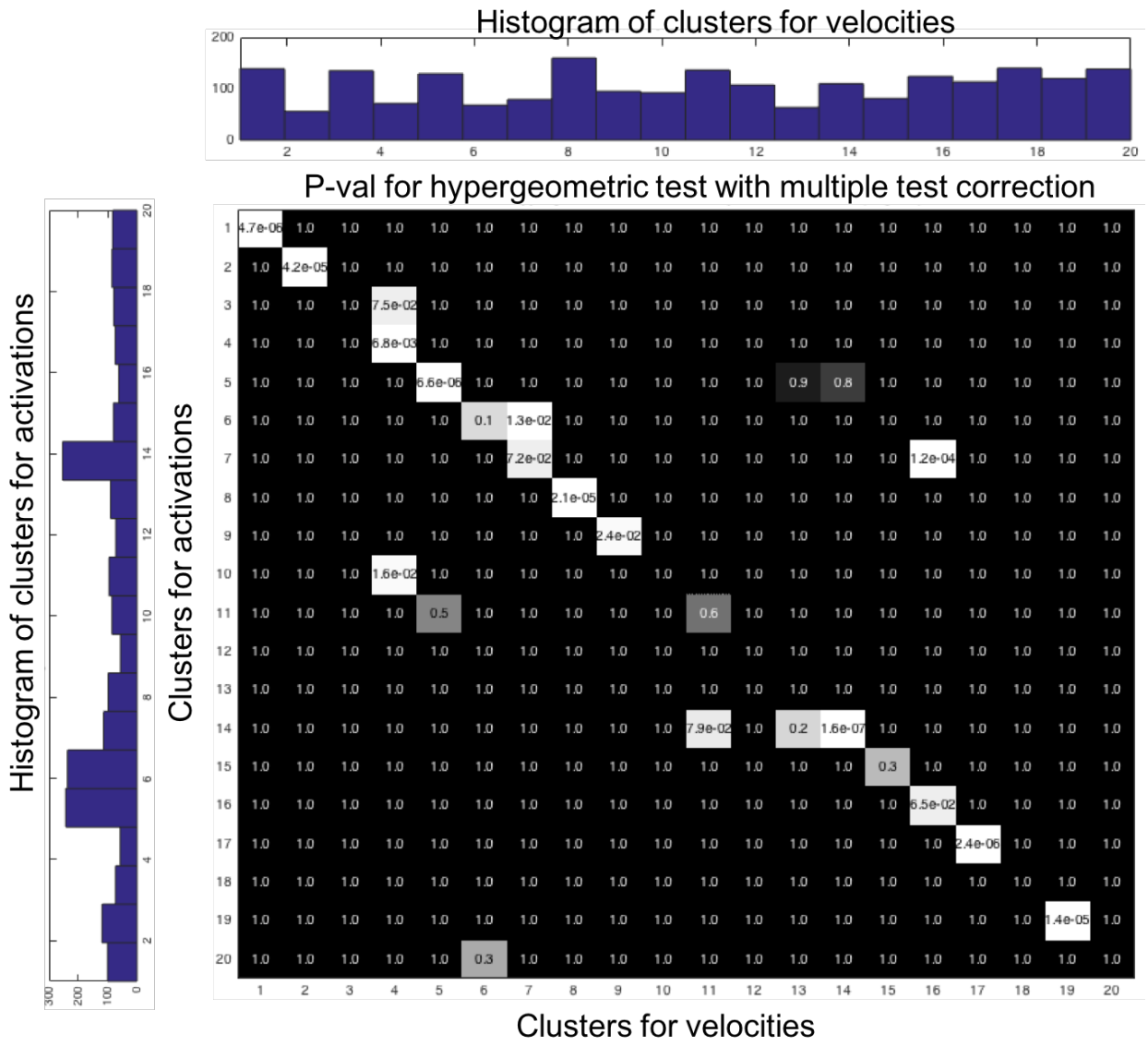


Figure 5.7: Hypergeometric testing results. The values indicate the p-values for hypergeometric testing for the enrichment of clusters based on clustering with GTPase activations and in clusters based on clustering with cell velocities(p-values are corrected using Bonferroni method for multiple tests (Shaffer, 1995)). The null hypothesis is that the clusters with cell velocities are not enriched in clusters with GTPase activations. The left and top blue bars show the distributions of number of data points in different clusters for activations and velocities respectively.

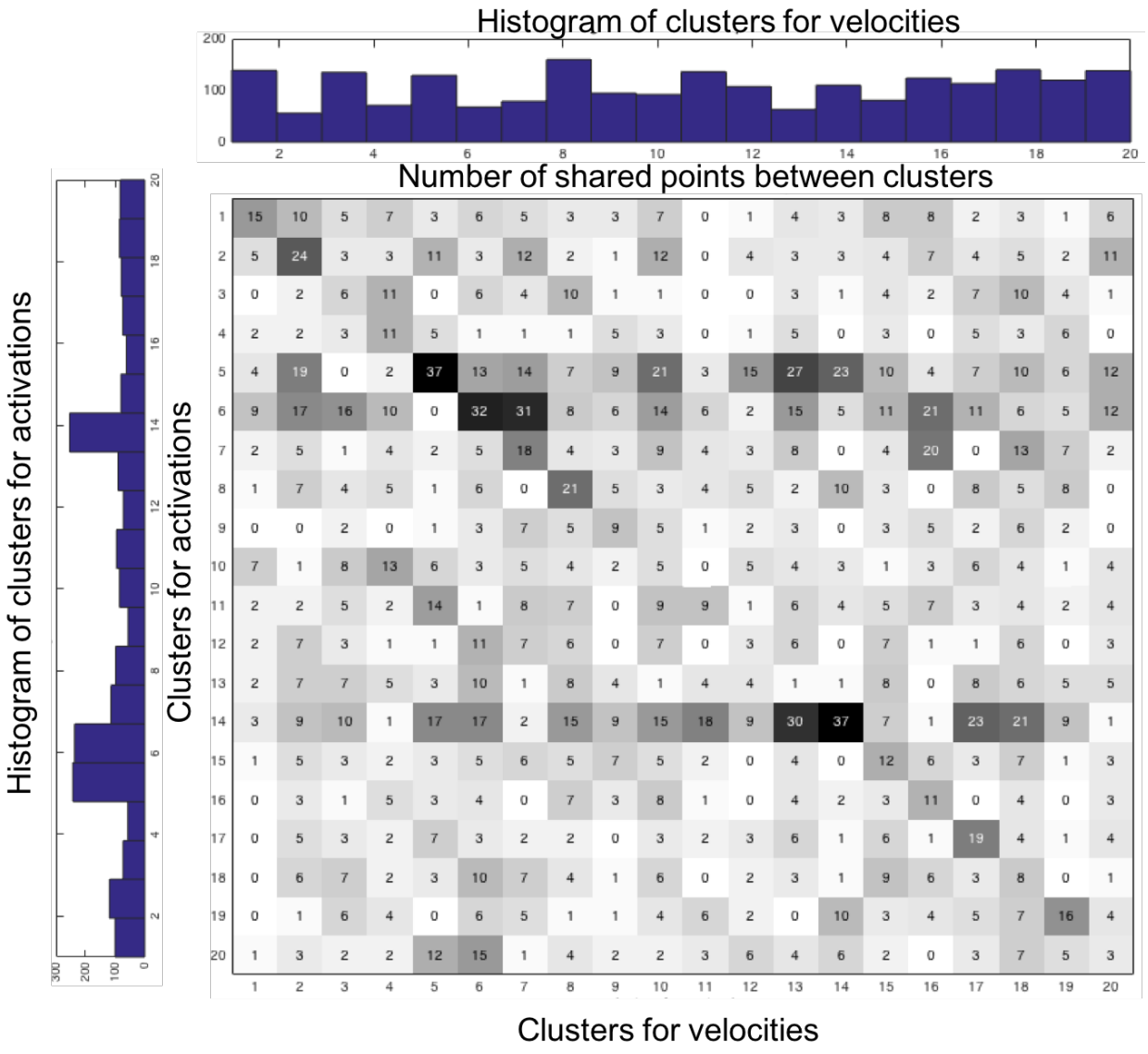


Figure 5.8: Intersection between different clusters for activations and velocities. The values indicate the number of intersected data points between different clusters for activations and velocities.

5.2.4 Modeling the relationship between Activations and Velocities based on Coupled Dictionary Learning

In Section 2.5, I introduced a coupled dictionary learning method to learn a joint basis for data from two coupled spaces. Then the data from two coupled spaces can be reconstructed as a linear combination of the basis vectors. These basis vectors can also be also considered as the common patterns for the data in activations and velocities. As a result, a coupled dictionary D is learned on concatenated data points for GTPase activations and cell velocities. Here, D is learned by solving the following dictionary learning problem,

$$\{\hat{D}, \hat{\alpha_i}\} = \underset{D,\alpha_i}{\operatorname{argmin}} \frac{1}{2} \sum_i \|x_i - D\alpha_i\| + \lambda \|\alpha_i\|_1, \tag{5.3}$$

where α is the coefficient for dictionary atoms to reconstruct the data x_i, x_i is the ith data point $x_i = [\mathbf{a}_{(c)i}^T \mathbf{v}_{(c)i}^T]^T$, where $\mathbf{a}_{(c)}$ and $\mathbf{v}_{(c)}$ indicate the activation and velocity patches over time respectively. The α_i give us a new representation of the data x_i based on dictionary atoms and can be used in prediction. Figure 5.9 shows an example for a data point represented as a linear combination of dictionary atoms.

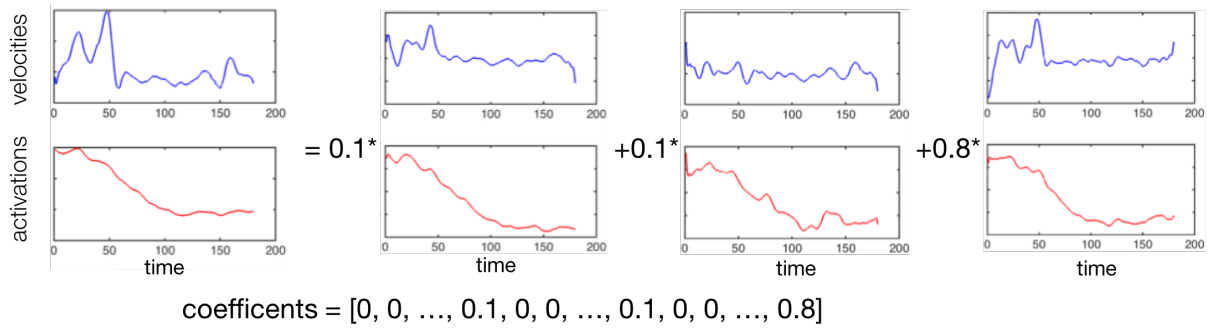


Figure 5.9: Example of a data point which is represented as linear combination of dictionary atoms. Blue and red curves indicates protrutions and activations respectively. The numbers are coefficients α.

5.2.5 Predicting Velocities from Activations based on Coupled Dictionary

Given a new boundary point with activations $a_{(c)i}$, it can be reconstructed by solving the sparse coding problem,

$$\{\hat{\alpha}_i\} = \underset{\alpha_i}{\text{argmin}} \frac{1}{2}\|a_{(c)i} - D_a\alpha_i\| + \lambda\|\alpha_i\|_1, \tag{5.4}$$

where D_a is the part of the coupled dictionary corresponding to activations. Thus the prediction of velocity v_i is,

$$v_{(c)i} = D_v\alpha_i, \tag{5.5}$$

where D_v is the part of the coupled dictionary corresponding to velocities.

5.3 Results

The prediction method is tested on the selected data. These data points are first extracted from 19 cell boundaries with equal distance, then the 11 neighboring data points are concatenated into a patch and each patch has an overlap of 5 data points with each other. About 2000 data patches are extracted from 19 cells. The data patches are selected based on the enrichment analysis described in the previous section, and 400 patches are left after selection.

Leave-one-out cross-validation is used to estimate generalization error. Each time the data points from one cell are held out for testing and the other data points are used for training. The prediction results based on coupled dictionary learning are compared with nearest neighbor search (NN) and PCA.

5.3.1 Prediction Results

Table 5.2 and Table 5.1 show the prediction errors using CDL, PCA and NN for the data with selection and without selection respectively. The prediction errors are larger for data without selection. This indicates that the selection strategy was able to identify data for which prediction is easier. However, for PCA and NN methods, the data without selection is more difficult to predict based on the prediction errors. Surprisingly, the prediction errors for data without selection are worse than the prediction errors for the trivial zero vector prediction. A possible reason is

Table 5.1: Prediction results using coupled dictionary learning (CDL), PCA and Nearest Neighbor search (NN) for GTPase activations(act) and velocities (vec) on all data. The numbers in the first row indicate the number of cluster centers and dictionary atoms. 'Dn' means CDL with only 'n' atoms. The prediction error is **3.32** if the trivial prediction (an all zero vector) is used.

D1	D10	D50	D100	D200	PCA	NN
3.95±0.49	3.84±0.43	3.85±0.51	3.82±0.46	**3.81±0.45**	4.31±0.81	4.28±0.98

Table 5.2: Prediction results using coupled dictionary learning (CDL), PCA and Nearest Neighbor search (NN) for GTPase activations(act) and velocities (vec) on selected data after enrichment analysis. The numbers in the first row indicate the number of cluster centers and dictionary atoms. 'Dn' means CDL with only 'n' atoms. The prediction error is **3.32** if the trivial prediction (an all zero vector) is used.

D1	D10	D50	D100	D200	PCA	NN
3.52±0.48	3.22±0.51	3.13±0.52	3.05±0.46	**3.02±0.41**	3.95±0.74	3.88±0.91

that most of the data involves complicated relationships or noise which cannot be captured by enrichment analysis (about 80%), thus prediction of these data points may not be feasible. The coupled dictionary learning based results achieve lower reconstruction errors for the selected data. I also notice that the performance saturates when the number of dictionary atoms is larger than 100 which means that 100 dictionary atoms are sufficient to explain the data in the coupled spaces. The data dimension is about 650.

Figure 5.10 shows an example of the prediction results. CDL based prediction can show a better prediction result compared with other methods. A discussion of the results of different methods is given in Section 5.3.1.1.

5.3.1.1 Discussion of Prediction Results

NN: The NN based method predicts the velocities of the test data by first finding the nearest neighbor of activations for the test data in the training samples. These corresponding velocities are used as the predictions.

This method works well when there are enough training samples which can cover the whole space or the subspace that the data lives in. However, in my test, only 400 data points are used for prediction which may not be enough to cover the whole space. Therefore, the test data's neighborhood may not be represented in the training data which results in an inaccurate prediction.

PCA: The PCA method is introduced in Appendix A.2. It aims at maximizing the variance of projected data along the principal components thus minimizing the reconstruction error between the original data and its estimate (Qin and Dunia, 2000). Given data $X = [x_1 \dots x_n] \in \mathbb{R}^{p \times n}$, and principal components $V_m = [v_1 \dots v_m] \in \mathbb{R}^{p \times m}$, the estimate of X is

$$\hat{X} = V_m^T \alpha, \tag{5.6}$$

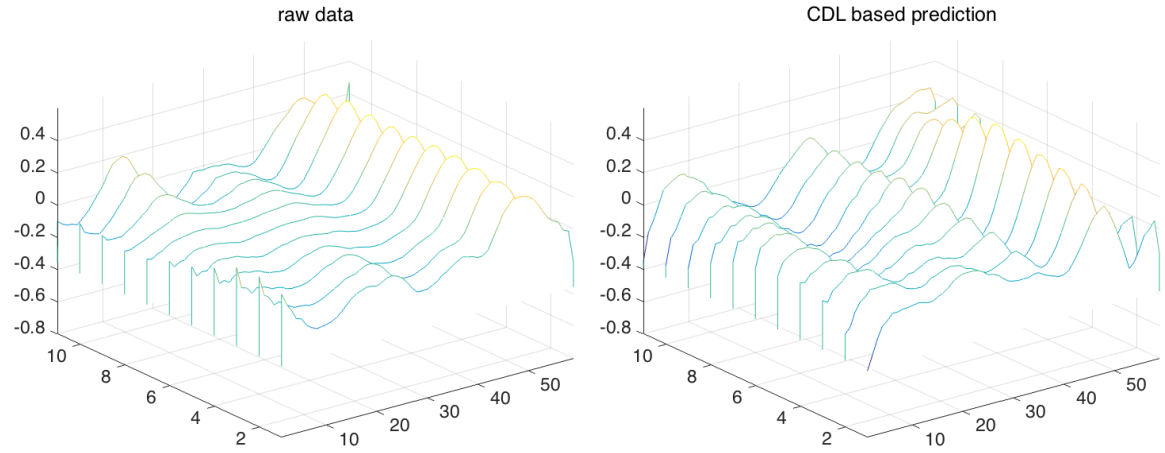
where $\alpha = V_m^T X$. The standard PCA only handles data in one space. In order to use PCA for two coupled spaces, I concatenate the data from two different spaces. Thus I have $\tilde{X} = [\tilde{x}_1, \dots, \tilde{x}_n]$, where $\tilde{x}_i = [a_i^T \ v_i^T]^T$ and the corresponding principal components $\tilde{V}_m$. To predict velocities based on activations, the first step is to estimate the α. However, as I only know a_i from the test data, the α cannot be obtained from $\tilde{V}_m^T \tilde{X}$. I can estimate the α by minimizing the reconstruction error for the activations

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \sum_i \|a_i - \tilde{V}_{(a)m} \alpha\|, \tag{5.7}$$

where $\tilde{V}_{(a)m}$ is the part of the principal components corresponding to activations. The velocities can be obtained from $\tilde{V}_{(v)m} \hat{\alpha}$ where $\tilde{V}_{(v)m}$ is the other part of the principal components corresponding to the velocities. There is a potential problem for this approach. α in Equation (5.6) is the projection of X on the orthogonal matrix V_m^T, while $\hat{\alpha}$ is the estimate from solving Equation (5.7). Because α is usually not $\hat{\alpha}$ which cannot be obtained by simple projections on the vectors of $\tilde{V}_{(a)m}$. Instead an equation system needs to be solved. Because α is usually not $\hat{\alpha}$, the $\hat{\alpha}$ coefficient which can minimize the reconstruction error for activations may not be a good estimate of the coefficient for velocities. There is no guarantee that α is equal to $\hat{\alpha}$. The result in Table 5.2 also shows that the prediction based on PCA is not very good compared to CDL.

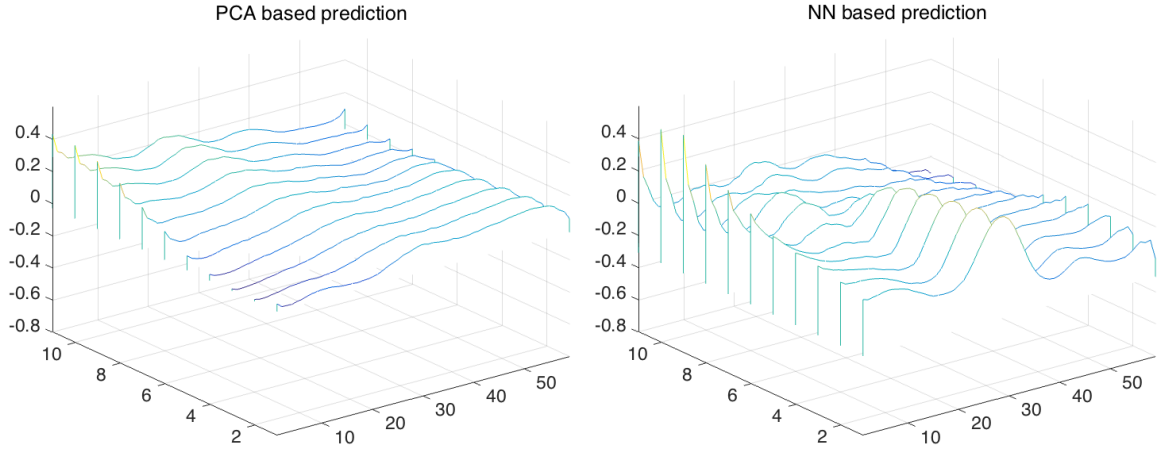
5.3.2 Common Patterns for Activations and Protrusions

A coupled dictionary is a set of basis vectors that can be used to represent the data from two coupled spaces. Figure 5.9 shows an example of the representation of data using dictionary atoms. Dictionary atoms can be considered as the common patterns for activations and protrusions.



(a) raw data point to predict

(b) CDL based prediction



(c) PCA based prediction

(d) NN based prediction

Figure 5.10: Example of prediction results of (a) velocity data for (b) CDL, (c) PCA and (d) NN methods. The data vectors are reshaped into 11×59 matrices where 59 is the number of time points of the boundary movement. The data matrices show the velocities for the boundary points in one patch.

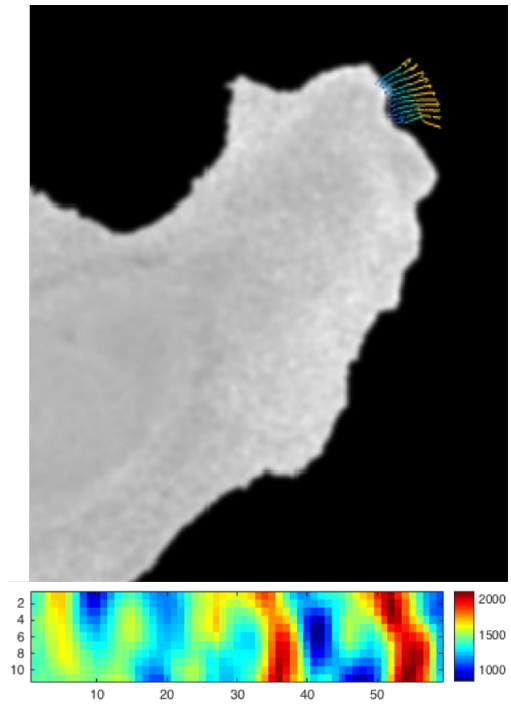
Figure 5.11 shows two selected patterns for activations and velocities and their corresponding cell regions. Here, the selected dictionary atoms are the dominant atoms used to explain the data points. Pattern 1 and pattern 2 in Figure 5.11 show cell protrusions and retractions respectively, and the corresponding dictionary atoms also show these patterns based on the displacement plots. The activations in pattern 1 and pattern 2 are also similar to the activations in the corresponding dictionary atoms which demonstrates that the dictionary atoms capture the pattern of activation and cell movements.

5.3.2.1 Relation to the Previous Work

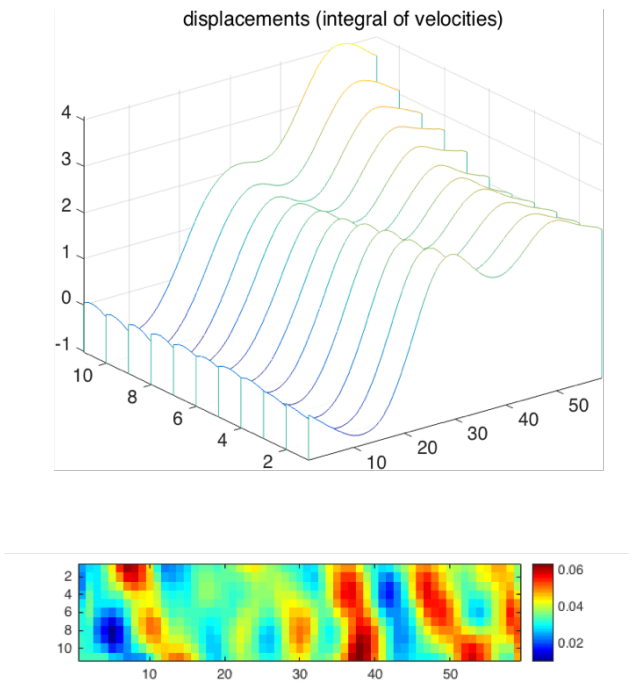
In (Machacek et al., 2009), the authors show that GTPase activations are correlated with cell movement with some time shifts, and these time shifts can be found using cross correlation. I also compute the cross correlation on the dictionary atoms from the learned coupled dictionary. The result of the cross correlation on rhoA activations and cell velocities is shown in Figure 5.12. The time delay for the activations with respect to velocities is about 10 seconds (time interval between two frames is 10 seconds), which is similar to the result in (Machacek et al., 2009) (6 seconds in paper). Note that the data I used to compute the cross correlation is different from the paper, but I can still obtain a similar result.

5.4 Discussion and Conclusion

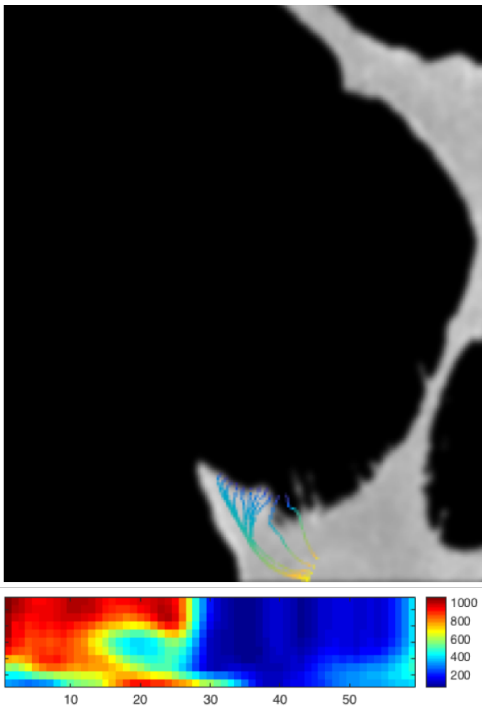
This chapter proposed methods for the analysis of the relationship between GTPase activations and cell protrusions. Enrichment analysis is first applied to the data clusters after k-means clustering on activations and protrusions. It provides information about whether one cluster or group of data points is enriched in other clusters or groups. If boundary points clustered together according to velocities, also cluster together according to activations, it means these clusters share some boundary points which suggests that the activations and protrusions for these points are coupled with each other. Then I investigated the predictability of velocities from activations by using coupled dictionary learning. Coupled dictionary based prediction achieved the best results compared to



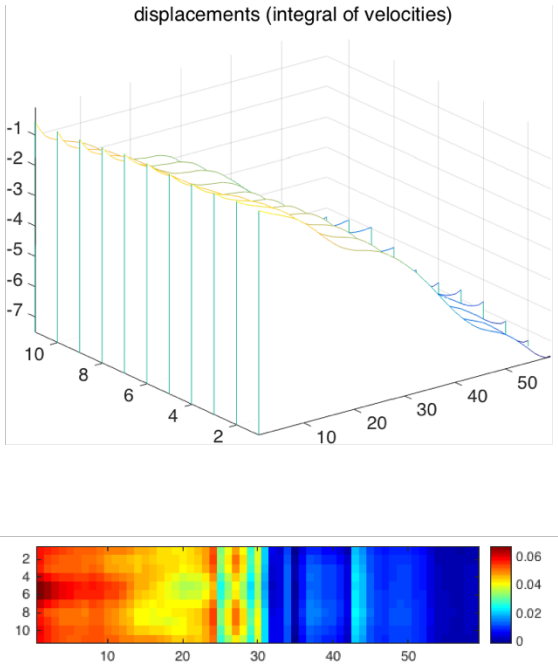
(a) cell region corresponding to pattern 1



(b) dictionary atom corresponding to pattern 1



(c) cell region corresponding to pattern 2



(d) dictionary atom corresponding to pattern 2

Figure 5.11: Example of activations and cell movement patterns. In (a) and (c), the traces of boundary points are superimposed on the cell images, the color of the line indicates time t where blue indicate $t = 0$. The map below each cell image is the activation map corresponding to these boundary points (y axis indicates the index of boundary points and x axis represents time). In (b) and (d), the dictionary atoms are shown to match the corresponding cell movement and activation patterns in (a) and (c) respectively. The displacement plots show the integral of velocities in dictionary atoms, the map below the displacement plot is the corresponding activation map of the dictionary atoms.

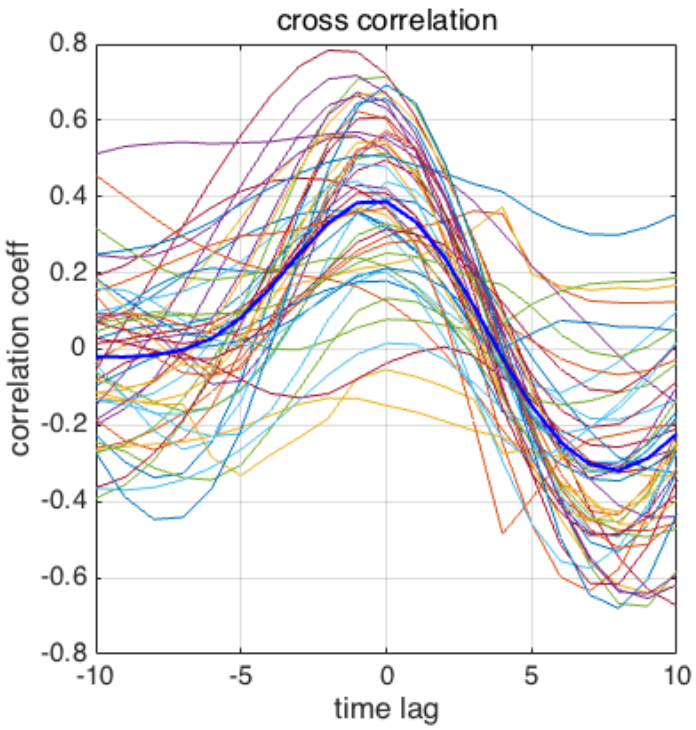


Figure 5.12: Cross correlation for dictionary atoms between rhoA activations and velocities. X axis indicates different lags between GTPase activations and cell velocities.

other tested methods. I also provided a detailed discussion on the prediction results obtained from other methods.

Future work includes testing the proposed method on a larger data set. The major problem for the current project is the lack of data. If I have more data, I could potentially find stronger common patterns for GTPase activations and cell velocities. Another future work could be real experiments to validate these patterns, i.e. by creating activation patterns in a cell, I could test if they indeed result in the predicted cell boundary changes.

CHAPTER 6: ROBUST COUPLED DICTIONARY LEARNING

In Chapter 3, I proposed a coupled dictionary learning method for multi-modal registration. The learned dictionary can be applied in a sparse representation framework to transform an input image from one modality to another. However, coupled dictionary learning may be impaired by lack of correspondence between image modalities in the training data, for example due to areas of low quality in one of the modalities. Such non-corresponding areas can negatively influence dictionary learning, thereby the joint representation of images, and consequentially the image analogy results. In this chapter, I propose a probabilistic model that explicitly accounts for image areas that are poorly corresponding between the image modalities. As a result, I can cast the problem of learning a dictionary in the presence of problematic image patches as a likelihood maximization problem and solve it with a variant of the EM algorithm (Dempster et al., 1977). The algorithm iterates identification of the poorly corresponding patches and refinements of the learned dictionary. I tested the method on both synthetic and real data which showing an improvement for image prediction problem compared to the standard coupled dictionary learning method.

6.1 Introduction

Coupled dictionary learning is challenging: it may fail or provide inferior dictionary quality without sufficient correspondences between modalities in the training data. This problem has so far not been addressed in the literature. For example, a low quality image deteriorated by noise in a modality can hardly match a high quality image in another modality. Furthermore, training images are pre-registered. Registration error may harm image correspondence and hence dictionary learning. Such noise- and correspondence- corrupted dictionaries will consequentially produce inferior results for image reconstruction or prediction. Figure 6.1 shows an example of coupled dictionary learning for both perfect and imperfect corresponding image pairs.

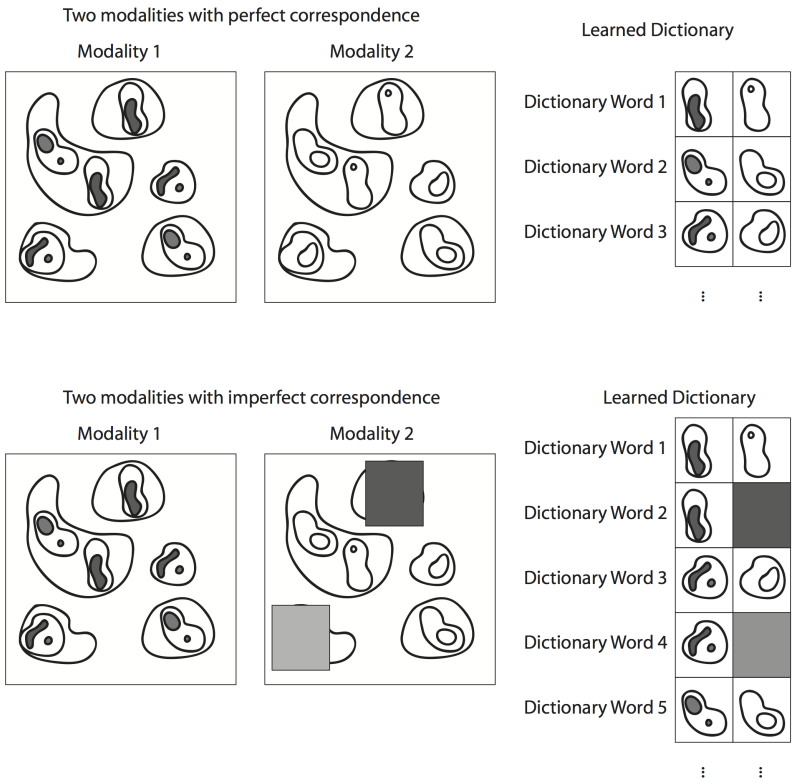


Figure 6.1: An illustration of perfect (left) and imperfect (right) correspondence between multi-modal images and their learned dictionaries. The imperfect correspondence (gray part in right images) could result in learning an imperfect dictionary (gray dictionary words) which is not desirable. Our goal is to *robustly* recover a compact dictionary of *corresponding* elements.

In this chapter, instead of directly learning a coupled dictionary from training data as in (Cao et al., 2012), I distinguish between image regions with and without good correspondence in the learning process. 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.

The main contributions are as follows,

- *I propose a probabilistic model for dictionary learning which discriminates between corresponding and non-corresponding patches.* This model is generally applicable to coupled dictionary learning.
- *I provide a method robust to noise and mis-correspondences.* I demonstrate this using real and synthetic data and obtain "cleaner" dictionaries.
- *I demonstrate consistency of performance for a wide range of parameter settings.* This indicates the practicality of the proposed approach.

This chapter is organized as follows: first, I describe the method and my probabilistic model in Section 6.2. I interpret the model in Section 6.3. In Section 6.4, I apply the model to both synthetic and real data. I also apply the proposed dictionary learning method to multi-modal image registration. This chapter concludes with a summary of results and an outlook on future work in Section 6.5.

6.2 Method

Let I_1 and I_2 be two different training images acquired from different modalities for the same area or object. Assume the two images have been registered already. A coupled dictionary can be learned from corresponding image patches extracted from I_1 and I_2.

6.2.1 Probabilistic framework for dictionary learning

In Section 2.4, I introduced the standard dictionary learning problem. Dictionary learning can also be interpreted in a probabilistic framework. The probabilistic model for dictionary learning

suggests that for each image patch x (Aharon et al., 2006),

$$x = D\alpha + \epsilon, \tag{6.1}$$

where ϵ is Gaussian white noise with zero mean and variance σ^2.

Assuming the reconstruction residual follows a Gaussian distribution, thus

$$\begin{aligned} p(x_i|D,\alpha_i) &= \mathcal{N}(x_i; D\alpha_i, \sigma^2) \propto \exp\{-\frac{1}{\sigma^2}\|x_i - D\alpha_i\|_2^2\}, \\ p(\alpha_i) &= \mathcal{L}(\alpha_i; 0, \lambda) \propto \exp\{-\lambda\|\alpha_i\|_1\}. \end{aligned} \tag{6.2}$$

Based on Bayes rule, I have the posterior $p(D|x_i) \propto p(x_i|D)p(D)$. Assuming the x_i to be independent, thus

$$p(x_i|D) = \int p(x_i, \alpha|D)d\alpha = \int p(x_i|D,\alpha)p(\alpha)d\alpha, \tag{6.3}$$

where α is a latent variable with Laplace distribution (Aharon et al., 2006). Based on Equation (6.3), Equation (6.2), and replacing the integration with the maximum of α (due to the high peak of the posterior in Equation (6.3) and Laplace prior of α) (Aharon et al., 2006), I obtain D by maximizing the log likelihood function,

$$\begin{aligned} \{\hat{D}, \hat{\alpha_i}\} &= \arg\max_D \sum_{i=1}^{N} \max_{\alpha_i} \{\log(p(x_i, \alpha_i|D))\} \\ &= \arg\max_D \sum_{i=1}^{N} \max_{\alpha_i} \{\log(p(x_i|D,\alpha_i)p(\alpha_i))\} \\ &= \arg\max_D \sum_{i=1}^{N} \max_{\alpha_i} \{\log(p(x_i|D,\alpha_i)) + \log(p(\alpha_i))\} \\ &= \arg\min_{\{D,\alpha_i\}} \sum_{i=1}^{N} \frac{1}{\sigma^2}\|x_i - D\alpha_i\|_2^2 + \lambda\|\alpha_i\|_1. \end{aligned} \tag{6.4}$$

Equation (6.4) can be solved with alternating optimization. First fixing D, I calculate the coefficients α_i, and then update the dictionary based on α_i. I update D and α iteratively until convergence is reached (Elad and Aharon, 2006).

6.2.2 Confidence measure for image patch

Before proposing my method, I first introduce the confidence measure for an image patch. 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 a good estimation because the "noisy" dictionary atoms are used in the reconstruction and vice versa.

Applying Bayes Rule (Mitchell, 1997; Besag, 1986), $p(h = 1|x_i)$ can be represented as,

$$p(h = 1|x_i) = \frac{p(x_i|h = 1)p(h = 1)}{p(x_i|h = 1)p(h = 1) + p(x_i|h = 0)p(h = 0)}. \tag{6.5}$$

Now the problem becomes how to estimate the likelihood $p(x_i|h)$ and prior $p(h)$. 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

$$\begin{aligned} p(x_i|h = 1, \theta_1) &= \mathcal{N}(x_i; \mu_1, \sigma_1^2), \\ p(x_i|h = 0, \theta_0; D, \alpha_i) &= \mathcal{N}(x_i - D\alpha_i; 0, \sigma_0^2). \end{aligned} \tag{6.6}$$

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

$$p(h = 1) = \pi, \text{ then } p(h = 0) = 1 - \pi. \tag{6.7}$$

Based on the assumption of conditional independence of the random variable x_i given h and θ (Gaussian Naive Bayes (GNB) assumption) (Mitchell, 1997), I can use either maximum likelihood estimates (MLE) or maximum a posteriori (MAP) estimates for these parameters. The MLE estimator for σ_k^2 is

$$\begin{aligned}\hat{\sigma}_0^2 &= \frac{1}{\sum_i \delta(h_i = 0)} \sum_i (x_i - \mu_i)^2 \delta(h_i = 0), \\ \hat{\sigma}_1^2 &= \frac{1}{\sum_i \delta(h_i = 1)} \sum_i (x_i - D\alpha_i)^2 \delta(h_i = 1),\end{aligned} \tag{6.8}$$

where the subscript i refers to the ith training sample, and the indicator function $\delta(h = 1)$ equals to 1 if $h = 1$ and 0 otherwise.

6.2.3 EM algorithm

The EM algorithm is a popular technique which can be used to solve maximum-likelihood parameter estimation problem in the presence of missing or hidden data(Dempster et al., 1977).

6.2.3.1 Maximum-likelihood

Suppose I have a density function $p(x|\theta)$ where θ indicates the set of parameters, for example, for a Gaussian distribution, θ could be means and covariances. For a training set of size N, $\mathbf{X} = \{x_i, \ldots, x_N\}$ which is independently drawn from this distribution, the density for the samples is,

$$p(\mathbf{X}|\theta) = \prod_{i=1}^{N} p(x_i|\theta) = L(\theta|\mathbf{X}).$$

The likelihood is a function of parameters θ given the data $\mathbf{X}$. Thus in the maximum likelihood problem, I find $\hat{\theta}$ such that,

$$\hat{\theta} = \underset{\theta}{\text{argmax}}\ L(\theta|\mathbf{X}).$$

Usually I minimize the log likelihood instead, thus I have,

$$\hat{\theta} = \underset{\theta}{\text{argmin}}\ -\log\ L(\theta|\mathbf{X}).$$

Since $\log(x)$ is a strictly increasing function, the value of θ that maximizes $p(\mathbf{X}|\theta)$ also maximizes $L(\theta)$. Explicitly estimating the parameters θ by solving the maximum likelihood problem may be hard depending on the form of $p(x|\theta)$.

6.2.3.2 EM algorithm

The EM algorithm gives an efficient iterative procedure to solve the maximum likelihood estimation problem by introducing some latent variables (Dempster et al., 1977; Redner and Walker, 1984). Maximizing $L(\theta)$ explicitly might be difficult. The strategy of the EM algorithm is to repeatedly construct a lower-bound of L (E-step), and then optimize the lower-bound (M-step). Suppose I have some latent random vector $\mathbf{Z}$. The total probability $p(\mathbf{X}|\theta)$ can be written as,

$$p(\mathbf{X}|\theta) = \sum_z p(\mathbf{X}|z,\theta)p(z|\theta)$$

Our objective is to maximize $L(\theta)$. Assume after the nth iteration the current estimate for θ is given by θ_n. I want to maximize the difference (Bishop, 2006),

$$\begin{aligned}
L(\theta) - L(\theta_n) &= \log\ p(\mathbf{X}|\theta) - \log\ p(\mathbf{X}|\theta_n) \\
&= \log \sum_z p(\mathbf{X}|z,\theta)p(z|\theta) - \log\ p(\mathbf{X}|\theta_n) \\
&= \log \sum_z p(\mathbf{X}|z,\theta)p(z|\theta).\frac{p(z|\mathbf{X},\theta_n)}{p(z|\mathbf{X},\theta_n)} - \log\ p(\mathbf{X}|\theta_n) \\
&= \log \sum_z p(z|\mathbf{X},\theta)(\frac{p(\mathbf{X}|z,\theta)p(z|\theta)}{p(z|\mathbf{X},\theta_n)}) - \log\ p(\mathbf{X}|\theta_n) \\
&\geq {}^{1} \sum_z p(z|\mathbf{X},\theta) \log(\frac{p(\mathbf{X}|z,\theta)p(z|\theta)}{p(z|\mathbf{X},\theta_n)}) - \log\ p(\mathbf{X}|\theta_n) \\
&= \sum_z p(z|\mathbf{X},\theta) \log(\frac{p(\mathbf{X}|z,\theta)p(z|\theta)}{p(z|\mathbf{X},\theta_n)p(\mathbf{X}|\theta_n)}) \\
&= \Delta(\theta|\theta_n).
\end{aligned} \tag{6.9}$$

[1] Using Jensen's inequality. Refer to appendix for more details.

I define $\ell(\theta|\theta_n) = L(\theta_n) + \Delta(\theta|\theta_n)$, and based on Equation (6.9), I have

$$L(\theta) \geq \ell(\theta|\theta_n). \tag{6.10}$$

From Equation (6.10), the $\ell(\theta|\theta_n)$ is the lower-bound of the likelihood function $L(\theta)$, and the value of functions $L(\theta)$ and $\ell(\theta|\theta_n)$ are equal when $\theta = \theta_n$. In order to maximize the value of $L(\theta)$, the EM algorithm selects θ that can maximize $\ell(\theta|\theta_n)$. Thus I have,

$$\begin{aligned}
\theta_{n+1} &= \underset{\theta}{\text{argmax}}\{\ell(\theta|\theta_n)\} \\
&= \underset{\theta}{\text{argmax}}\{L(\theta_n) + \sum_z p(z|\mathbf{X},\theta)\log(\frac{p(\mathbf{X}|z,\theta)p(z|\theta)}{p(z|\mathbf{X},\theta_n)p(\mathbf{X}|\theta_n)})\} \\
&\text{drop } L(\theta_n) \text{ which is constant with respect to } \theta \\
&= \underset{\theta}{\text{argmax}}\{\sum_z p(z|\mathbf{X},\theta)\log(p(\mathbf{X}|z,\theta)p(z|\theta))\} \\
&= \underset{\theta}{\text{argmax}}\{\sum_z p(z|\mathbf{X},\theta)\log(p(\mathbf{X},z|\theta))\} \\
&= \underset{\theta}{\text{argmax}}\{E_{z|\mathbf{X},\theta}[\log\, p(\mathbf{X},z|\theta)]\}
\end{aligned} \tag{6.11}$$

From Equation (6.11), I have the EM algorithm:

1. **E-step** : Determine the expectation $E_{z|\mathbf{X},\theta}[\log\ p(\mathbf{X},z|\theta)]$;
2. **M-step** : Maximize the expectation with respect to θ.

6.2.4 Robust coupled dictionary learning based on EM algorithm

For robust coupled dictionary learning, I want to estimate $\theta = \{\tilde{D}, \alpha\}$ considering the latent variable h. Based on the probabilistic framework of dictionary learning (Aharon et al., 2006), I have $p(\tilde{x}|\theta) = \sum_h p(\tilde{x},h|\theta)$.

The ML estimation for θ is as follows

$$\hat{\theta} = \arg\max_\theta p(\tilde{x}|\theta) = \arg\max_\theta \log\sum_h p(\tilde{x},h|\theta). \tag{6.12}$$

Let $\ell(\theta) = \log \sum_h p(\tilde{x}, h|\theta)$. Instead of directly maximizing $\ell(\theta)$, I maximize the lower bound $Q(\theta) = \sum_h p(h|\tilde{x}, \theta) \log p(\tilde{x}, h|\theta)$ (Neal and Hinton, 1998). $p(h|\tilde{x}, \theta)$ is the confidence in section 6.2.2. I can apply the following EM algorithm to maximize $Q(\theta)$,

$$\begin{aligned} &\textbf{E-step} : Q(\theta|\theta^{(t)}) = E_{h|\tilde{x},\theta}[\log p(\tilde{x}, h|\theta^{(t)})]; \\ &\textbf{M-step} : \theta^{(t+1)} = \arg\max_{\theta} Q(\theta|\theta^{(t)}). \end{aligned} \tag{6.13}$$

Algorithm 6 EM algorithm for Coupled Dictionary Learning

Input: Training Coupled image patches: $\{\tilde{x}_i\}$, $i \in 1, ..., N$;
Initialize Coupled dictionary $\tilde{D} = \tilde{D}_0$, $\tilde{D}_0$ is trained on all of the $\tilde{x}_i$;

Output: Refined dictionary $\hat{\tilde{D}}$

1: (**E-step**) determine $E_{h|\tilde{x},\theta}[\log p(\tilde{x}, h|\theta^t)] = \sum_h \delta_p(p(h = 0|\tilde{x}_i, \theta^t)) \log p(\tilde{x}, h|\theta^t)$
by computing $\delta_p(p(h = 0|\tilde{x}_i, \theta^t))$, where (Besag, 1986; Neal and Hinton, 1998)

$$\delta_p(p) = \begin{cases} 1, \text{ if } p \geq 0.5, \\ 0, \text{ otherwise.} \end{cases} \tag{6.14}$$

$$p(h = 0|\tilde{x}_i, \theta) = \frac{p(\tilde{x}_i|h = 0, \theta)p(h = 0)}{p(\tilde{x}_i|h = 1, \theta)p(h = 1) + p(\tilde{x}_i|h = 0, \theta)p(h = 0)}. \tag{6.15}$$

2: (**M-step**) maximize $E_{h|\tilde{x},\theta}[\log p(\tilde{x}, h|\theta^t)]$
by updating $\tilde{D}$ and α as follows,

$$\begin{aligned} \tilde{D}^{(t+1)} &= \arg\min_{\tilde{D}} \sum_{i=1}^{N} \delta_p(p(h = 0|\tilde{x}_i, \theta^t))(\frac{1}{2}\|\tilde{x}_i - \tilde{D}\alpha_i^t\|_2^2 + \lambda\|\alpha_i^t\|_1), \\ &\text{s.t. } \|\tilde{D}_j\|_2^2 \leq 1, \ j = 1, 2, ..., k. \\ \alpha_i^{(t+1)} &= \arg\min_{\alpha_i} \delta_p(p(h = 0|\tilde{x}_i, \theta^t))(\frac{1}{2}\|\tilde{x}_i - \tilde{D}^{(t+1)}\alpha_i\|_2^2 + \lambda\|\alpha_i\|_1). \end{aligned} \tag{6.16}$$

3: Iterate E and M steps until convergence reached.

In the E-step I compute $p(h_i|\tilde{x}, \theta), \ h_i \in \{1, 0\}$ which provides a confidence for each training patch given $\tilde{D}$ and α. In the M-step $p(h_i|\tilde{x}, \theta)$ corresponds to a weight for each image patch for updating θ. In this chapter, I use a variant of EM algorithm for coupled dictionary learning. I use $\delta_p(p(h_i|\tilde{x}, \theta))$ replacing $p(h_i|\tilde{x}, \theta)$. Here, $\delta_p(p)$ is an indicator function and $\delta_p(p) = 1, \text{ if } p \leq 0.5, \ \delta_p(p) = 0, \text{ otherwise}$. Thus in each iteration I rule out the image patches which have high

confidence that they are based on a reconstruction with non-corresponding coupled dictionary items, and then refine the coupled dictionary using the corresponding training samples. The detailed algorithm is shown in Algorithm 6.

6.3 Interpreting the model

In Section 6.3.1, I discuss the dictionary learning as a generative model. In Section 6.3.2, I introduce a graphical model to interpret our method.

6.3.1 Generative model for dictionary learning

Based on the probabilistic framework for dictionary learning in Section 6.2.1, I can construct a graphical model to interpret the generative learning process of our proposed algorithm. The graphical model is shown in Figure 6.2.

Here the coupled dictionary $\tilde{D}$ is the parameter I want to estimate for the problem. h_i are latent variables and $\alpha_i = [\alpha_{1i}, \alpha_{2i}]^T$ has a Laplace prior (Here, α_{1i} and α_{2i} are corresponding to two modalities). h_i is a hypothesis whether the multi-modal image patch $\tilde{x}$ is generated by $\tilde{D}$ or by noise at mean $\{\mu_{1i}\mathbf{1}, \mu_{2i}\mathbf{1}\}$. Therefore $p(h_i = 1) = \pi$ is the prior of the noise level about the training patches. When $h_i = 0$, x_i are generated with a Gaussian probability distribution given $\tilde{D}$ and α_i, $p(\tilde{x}_i|\tilde{D}, \alpha_i) = p(\tilde{x}_i|h = 0, D, \alpha_i) \propto \exp(\frac{1}{\sigma_0^2}\|\tilde{x}_i - D\alpha_i\|_2^2)$, and for $h = 1$, x_i is generated with a Gaussian distribution given $\mu_i\mathbf{1}$, $p(\tilde{x}_i|\mu_i\mathbf{1}) = p(\tilde{x}_i|h = 1, \mu_i\mathbf{1}) \propto \exp(\frac{1}{\sigma_1^2}\|\tilde{x}_i - \mu_i\mathbf{1}\|_2^2)$.

Now the maximum likelihood estimation of $\tilde{D}$ for our generative model according to the marginal joint distribution

$$\begin{aligned} p(\{\tilde{x}_i\}_{i=1}^N, \tilde{D}, \alpha) &= \sum_j p(\{\tilde{x}_i\}_{i=1}^N, h_j, \tilde{D}, \alpha) \\ &\propto \sum_j p(\{\tilde{x}_i\}_{i=1}^N, h_j|\tilde{D}, \alpha) \end{aligned} \tag{6.17}$$

is exactly the same as the one what I discussed in section Section 6.2.4.

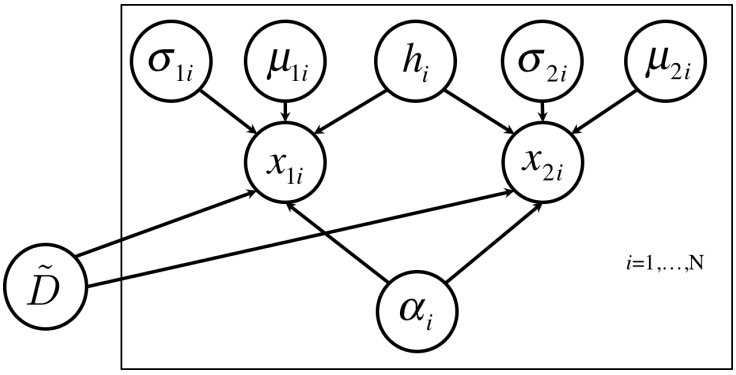


Figure 6.2: Graphical model for the proposed generative learning framework

6.3.2 Criterion for corresponding multi-modal image patch

If there is no prior information about $p(h)$, I assume $p(h = 1) = p(h = 0) = 0.5$. Based on (6.6), (6.14), (6.15), I have

$$\begin{aligned} p(h = 0|\tilde{x}_i, \theta) \geq 0.5 \Leftrightarrow \\ p(\tilde{x}_i|h = 0, \theta) \geq p(\tilde{x}_i|h = 1, \theta) \Leftrightarrow \\ \frac{1}{\sigma_0^2}\|\tilde{x}_i - \tilde{D}\alpha\|_2^2 \leq \frac{1}{\sigma_1^2}\|\tilde{x}_i - \mu_i\mathbf{1}\|_2^2 \Leftrightarrow \\ \|\tilde{x}_i - \tilde{D}\alpha\|_2^2 \leq \frac{\sigma_0^2}{\sigma_1^2}\|\tilde{x}_i - \mu_i\mathbf{1}\|_2^2 = c\|\tilde{x}_i - \mu_i\mathbf{1}\|_2^2. \end{aligned} \tag{6.18}$$

Here $\|\tilde{x}_i - \tilde{D}\alpha\|_2^2$ is the sum of squares of reconstruction residuals of image patch $\tilde{x}_i$, and $\|\tilde{x}_i - \mu_i\mathbf{1}\|_2^2$ is the sum of squares of centered intensity value (removed mean $\mu_i\mathbf{1}$) in $\tilde{x}_i$.

Thus (Equation (6.18)) defines a criterion for corresponding multi-modal image patches as those that can be explained by the coupled dictionary $\tilde{D}$ better than by the patch's mean intensity, i.e. the sum of squared residuals should be smaller than a threshold T, and T is dependent on the variance of $\tilde{x}_i$, σ_1^2 and the variances of reconstruction residual σ_0^2.

I first initialize $\sigma_0 = \sigma_1$ at the beginning of the EM algorithm. I can update σ_0 from reconstruction residuals of training data during the iterations, then the only parameter I need to choose is σ_1. Intuitively, a small σ_1 is in favor of more corresponding image patches and a large σ_1 considers more image patches as non-corresponding. Here, σ_1 can be considered as a direct estimation of the

noise level in the training set. I will show how different σ_1 affect my iterative dictionary learning and further the reconstruction results in Section 6.4.

6.4 Experimental validation

The image prediction problem (for a known dictionary $\tilde{D}$) amounts to solving

$$\{\hat{\alpha}_i\} = \arg\min_{\alpha_i} \sum_{i}^{N} \|\tilde{x}'_i - \tilde{D}\alpha_i\|_2^2 + \lambda\|\alpha_i\|_1. \tag{6.19}$$

where $\tilde{x}'_i = R_i[I_1, u_2]^T$ and u_2 is the prediction of I_2. Since I_2 is not measured, I can effectively set $R_i u_2 = D_2\alpha_i$ or equivalently remove it from the optimization. Given $\{\hat{\alpha}_i\}$, I can then compute the predicted image. Most applications using coupled dictionaries are concerned with the prediction residuals, such as super-resolution and multi-modal registration (Yang et al., 2010; Cao et al., 2012). I therefore first validate the algorithm based on the resulting sum of squared prediction residuals (SSR).

I test the proposed coupled dictionary learning method on synthetic and real data. For the synthetic data, I generate non-corresponding multi-modal image patches using the generative model in Section 6.3.1. First I choose $p(h = 1)$ which defines the noise level in the training set, i.e. the percentage of non-corresponding multi-modal image patches in the training set. For each non-corresponding patch x_i^1, I generate $\mu_i\mathbf{1}$ as the mean of all training patches and add Gaussian noise ϵ_μ. Finally, I generate a noise patch by adding Gaussian noise $\epsilon_{x_i^1}$ to the mean $\mu_i\mathbf{1}$. Then I have $x_i^1 = \mu_i\mathbf{1} + \epsilon_{x_i^1}$.

6.4.1 Synthetic experiment on textures

I create multi-modal textures by smoothing a given texture with a Gaussian kernel and inverting the intensity of the smoothed image. Thus I have corresponding textures from different 'modalities'. Figure 6.3 shows an example of my generated multi-modal textures. I generate both training and testing multi-modal textures from Figure 6.3, i.e. use half of the multi-modal textures for training (add noise as non-correspondence regions) and the other half of the multi-modal textures for testing.

I extract 10×10 image patches in both training images, and add 'noise' with non-corresponding image patches to replace corresponding patches. The σ for the Gaussian noise is set to 0.2.

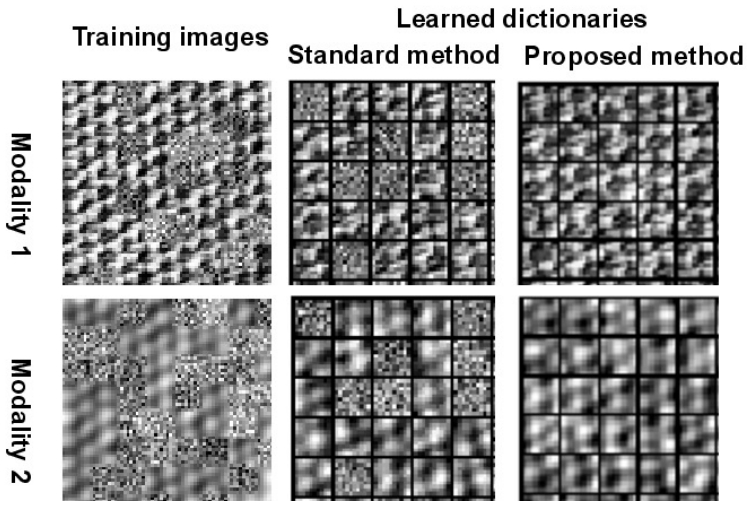
I test how σ_1 influences my dictionary learning method at a fixed noise level $p(h = 1) = 0.5$. Fig. 6.3 shows the result. In practice, I can either learn σ_1 with an EM algorithm or manually choose it. When σ_1 is close to 0.2 (the σ for the noise), to be specific, $\sigma_1 \in (0.15, 0.4)$, I get consistently lower SSRs. This indicates that the proposed algorithm is robust for a wide range of σ_1 values and noise. For $\sigma_1 < 0.15$, all the patches are considered as corresponding patches while for $\sigma_1 > 0.4$, all the patches are classified as non-corresponding patches. The method has the same performance as the standard method in (Cao et al., 2012) in these two cases. The learned coupled dictionaries are illustrated in Fig. 6.3 showing that the proposed algorithm successfully removes non-corresponding patches.

6.4.2 Synthetic experiment on multi-modal microscope images

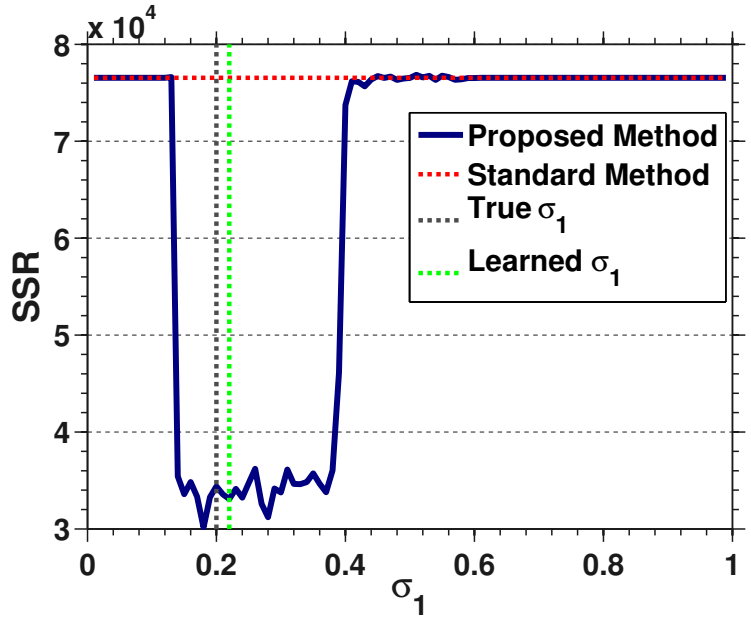
I also test the proposed algorithm on correlative microscope images. I have 8 pairs of Scanning Electron Microscopy (SEM) and confocal images. Image pairs have been aligned with fiducials. Figure 6.4 (a) illustrates an example of SEM/confocal images. I add non-corresponding patches using the same method as in Section 6.4.1. Figure 6.4 (a) shows the results. The dictionary learned with the proposed method shows better structure and less noise compared with the standard dictionary learning method. Figure 6.4 (b) shows the interaction between σ_1 and SSR with fixed $p(h = 1) = 0.5$. For $\sigma_1 < 0.16$, all the image patches are categorized as corresponding patches while for $\sigma_1 > 0.6$, all the patches are classified as non-corresponding patches. The proposed method has the same performance as the standard method under these conditions. I observe a large range of σ_1 values resulting in improved reconstruction results indicating robustness.

6.4.3 Multi-modal registration on correlative microscopy

I applied the proposed method for multi-modal registration (Cao et al., 2012). The multi-modal image registration problem simplifies to a mono-modal one using the coupled dictionary in a sparse representation framework. The test data is Transmission Electron Microscopy (TEM) and confocal

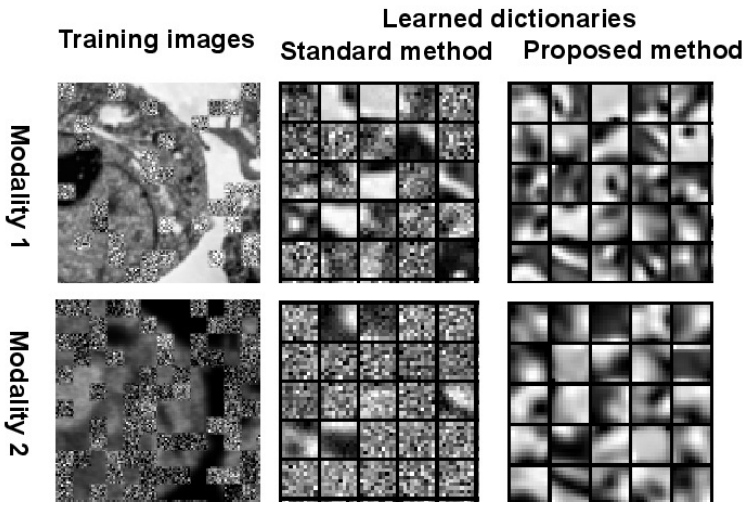


(a) training textures and learned $\tilde{D}$

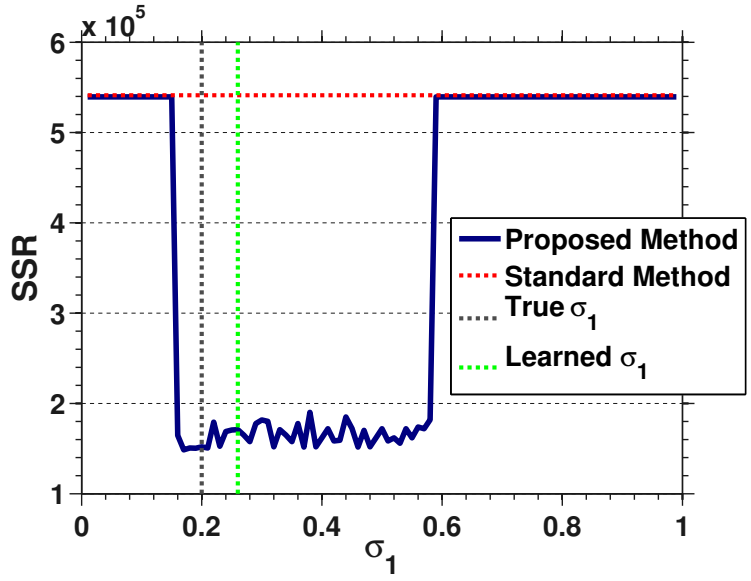


(b) SSR vs σ^1

Figure 6.3: $\tilde{D}$ is learned from training images with Gaussian noise (top). Standard method cannot distinguish corresponding patches and non-corresponding patches while our proposed method can remove non-corresponding patches in the dictionary learning process. The curve (bottom) shows the robustness with respect to σ_1. The vertical green dashed line indicates the learned σ_1.

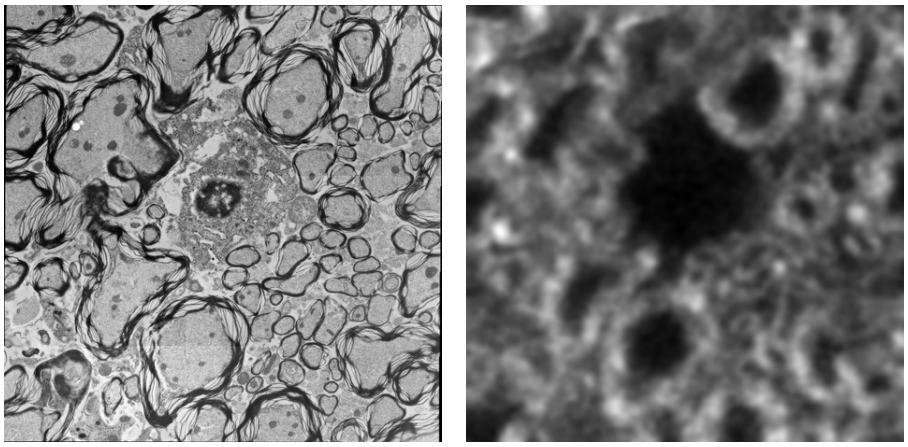


(a) training images and learned $\tilde{D}$



(b) SSR vs σ^1

Figure 6.4: $\tilde{D}$ is learned from training SEM/confocal images with Gaussian noise (top). The curve (bottom) shows the robustness with respect to σ_1. The vertical green dashed line indicates the learned σ_1.



(a) TEM (b) Confocal

Figure 6.5: TEM/Confocal images

microscopy which have been described in Section 3.4.1. I have six pairs of TEM/confocal images. I train the coupled dictionary using leave-one-out cross-validation. Figure 6.5 shows an example of our test data. I first registered the training images with manually chosen landmarks (no ground truth available), then learned the coupled dictionary and applied it to predict the corresponding image for a given source image. I resampled the predicted images with up to $\pm 2.07 \mu m$ (30 pixels) in translation in the x and y directions (at steps of 10 pixels) and $\pm 20^{\circ}$ in rotation (at steps of 10 degrees). Then I registered the resampled predicted image to the corresponding target using a rigid transformation model. σ_1 is chosen as 0.15 based on cross-validation for the prediction errors in this experiment. Table 6.1 shows a comparison of our method with the method in (Cao et al., 2012). The result shows about 15% improvement in prediction error and a statistically significant improvement in registration errors.

6.5 Conclusion

In this chapter, I proposed a robust coupled dictionary learning method based on a probabilistic formulation. I directly model corresponding and non-corresponding multi-modal training patches. The proposed method is based on a variant of the EM algorithm which classifies the non-corresponding image patches and updates the coupled dictionary iteratively. I validated the method using synthetic and real data. The proposed algorithm demonstrated its robustness to noise

Table 6.1: Prediction and registration results. Prediction is based on the method in (Cao et al., 2012), and I use SSR to evaluate the prediction results. Here, MD denotes my proposed coupled dictionary learning method and ST denotes the dictionary learning method in (Cao et al., 2012). The registrations use Sum of Squared Differences (SSD) and mutual information (MI) similarity measures. I report the results of mean and standard deviation of the absolute error of corresponding landmarks in micron (0.069 micron = 1 pixel). The p-value is computed using a paired t-test.

	Metric	Method	mean	std	p-value
Prediction	SSR	MD	$\mathbf{6.28 \times 10^4}$	3.61×10^3	$<$0.0001
		ST	7.43×10^4	4.72×10^3	
Registration	SSD	MD	**0.760**	0.124	0.0004
		ST	0.801	0.139	
	MI	MD	**0.754**	0.127	0.0005
		ST	0.795	0.140	

(non-corresponding image patches). I also applied the method to multi-modal registration showing an improvement in alignment accuracy compared with the traditional dictionary learning method. The proposed method is expected to be of general use for coupled dictionary learning. While the proposed method is based on a Gaussian noise model, it can easily be adapted to other noise models such as Poisson noise.

CHAPTER 7: DISCUSSION

This chapter revisits the thesis statement and contributions of this dissertation and discusses possibilities of future work.

7.1 Summary of contributions

This section lists each claim of contributions presented in chapter 1, followed by a discussion that relates to the claim and summarizes how it is addressed in this dissertation.

1. *I propose a sparse representation and coupled dictionary learning-based image analogy method to convert and thereby simplify a multi-modal registration problem to a mono-modal one.*

   The proposed method in Section 3.3 provides a general framework to synthesize the corresponding images of the input images based on example image pairs. The relationship between example image pairs can be captured by a coupled dictionary. Then the input images are reconstructed by the dictionary atoms corresponding to the modality of the input images, at the same time, the synthesized images for the other modality are estimated by the corresponding dictionary atoms for the other modality with the same coefficients (weights) in reconstruction.

   The major difference between the proposed method and the standard image analogy method is that the proposed method uses a sparse linear combination of dictionary atoms to predict each image patch instead of using a single nearest neighbor patch in standard IA. Also the proposed method learns a coupled dictionary from training patches thus can compress the search space (the number of dictionary atoms is much less than the training patches).

The proposed method converts image modalities from one to another, thereby simplifying the multi-modal registration problem to a mono-modal one. The proposed method is evaluated in Section 3.4 and shows its benefit for the registration between correlative microscope images.

The proposed method is essentially learning a customized distance measure based on the training data thereby better capturing differences between multi-modal images in registration.

2. *I propose a general framework for deformation estimation using appearance information based on coupled dictionary learning.*

   A general framework for deformation estimation based on the differences between a reference image and subject images is proposed in Chapter 4. The proposed method learns a semi-coupled dictionary for the difference image and deformation parameters. The goal of this project is to investigate the possibility of learning a coupled basis for difference image and deformation parameters.

   Different deformation parametrization methods are also studied in this section. As the difference image and deformation parameters form two spaces with large difference, a semi-coupled dictionary learning method is introduced to allow flexible coupling between dictionaries for two spaces. The method is evaluated in Section 4.4 on synthetic and real data for b-spline and initial momentum parametrization methods. The results show the performance improvements for the proposed method.

3. *I propose a framework for relating the spatio-temporal patterns between GTPase activations and cell movements of mouse embryonic fibroblasts (MEFs) based on coupled dictionary learning.*

   In Section 5.2.3, I use enrichment analysis to analyze the spatio-temporal relationship between GTPase activities and cell movements. The results suggest that predicting velocities based on GTPase activations is possible. I propose a coupled dictionary learning based prediction method in Section 5.2.4. The dictionary atoms can be considered as

common patterns for GTPase activations and cell velocities. The results in Section 5.3 show the examples of common patterns and their connection to real data.

4. *I develop a robust coupled dictionary learning method based on a probabilistic model which discriminates between corresponding and non-corresponding patches automatically.* The standard coupled dictionary learning method may fail due to non-corresponding image patches in the training data. In Chapter 6, I propose a robust coupled dictionary method to solve this problem. I introduce a latent variable to indicate whether an image patch is reconstructed by clean or noisy dictionary atoms during the learning process. As a result, I propose a probabilistic model for the coupled dictionary learning which can classify corresponding and non-corresponding patches. The model can be solved by an EM algorithm which is presented in Section 6.2.4.

   The proposed method is evaluated in Section 6.4 on both synthetic data and real data. The dictionary learning results show that the proposed method can successfully learn a clean dictionary by ruling out non-corresponding patches.

Having addressed the contributions above, I present again the thesis statement.

Thesis statement: 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 spatio-temporal 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.

7.2 Future Work

This section reviews possible directions of future research related to this dissertation.

7.2.1 Confidence in multi-modal registration

The coupled dictionary learning based image analogy usually cannot get perfect image synthesis results. Therefore the reconstruction error can be used to estimate the local confidence of the image

analogy result (while the prediction error is not known). If the reconstruction error is large which means that the local region is difficult to reconstruct the confidence of the image analogy result is low. This confidence estimate could then be used to weight the registration similarity metrics to focus on regions of high confidence.

7.2.2 Validation on large dataset from different modalities

The proposed method for multi-modal registration in Chapter 3 is only tested on a limited dataset. To further evaluate the benefit of the proposed method, a larger dataset is required. Also tests on different modalities would be interesting to see the limits of the proposed method.

7.2.3 Iterative deformation estimation

The proposed deformation estimation method is a one-pass algorithm, i.e. after learning the dictionary, the deformation is estimated in one run of the algorithm. However, there is no guarantee that the estimated result is good enough. It is possible to do this in an iterative approach, i.e. each time after obtaining an estimated deformation, I transform the image based on the obtained deformation and use the deformed image as an input of the algorithm. The process continues until converge.

7.2.4 Validation of cell velocities prediction on real cell data

The common patterns of GTPase activities and cell velocities can be captured by the learned coupled dictionary. However, the method is only validated on a limited dataset. It would be useful to do real experiments to validate these patterns, i.e. by creating activation patterns in a cell, I could test if they indeed result in the predicted cell boundary changes.

7.2.5 Robust dictionary learning for other noise models

The current robust dictionary learning method is based on the Gaussian noise model. Although I validated it on real data where the noise model may not be Gaussian, it would be nice to directly model the noise in the training data. The framework would be the same for different noise models, the only difference is how to compute the confidence in the EM algorithm.

APPENDIX A: APPENDIX FOR CHAPTER 5

A.1 Boundary Tracking

For a marker $p_i = (x_{p_i}, y_{p_i})$ on boundary Γ_t, $\nabla\phi_t(x_p, y_p)$ is the gradient of the level set ϕ_t at point p_i, thus the direction of $\nabla\phi_t(x_p, y_p)$ points in the direction with greatest increase of the distance of ϕ_t. Let $D_{t,t+1}(p_i)$ represent the distance from p_i on Γ_t to the closest point on Γ_{t+1}, $D_{t,t+1}(p_i) = -\phi_{t+1}(x_{p_i}, y_{p_i})$. If $D_{t,t+1}(p_i)$ is positive, there is a boundary protrusion from t to $t+1$ (the boundary point p_i is inside of the cell at time t+1), thus the search direction should point in the gradient direction $\nabla\phi_t(x_p, y_p)$ (based on the definition of protrusion and retraction). Similarly, if $D_{t,t+1}(p_i)$ is negative, there is a boundary retraction and the search direction should point in the negative gradient direction $-\nabla\phi_t(x_p, y_p)$. Thus the search distance for p_i at time t is defined as follows,

$$S_{p_i} = (D_{t,t+1}(p_i)\frac{\nabla\phi_{t+1}(x_{p_i}, y_{p_i})}{|\nabla\phi_{t+1}(x_{p_i}, y_{p_i})|}) \cdot \frac{\nabla\phi_t(x_{p_i}, y_{p_i})}{|\nabla\phi_t(x_{p_i}, y_{p_i})|}, \tag{A.1}$$

where '$\cdot$' represents the inner product operator of two vectors, p_i is the tracked ith boundary point at time t, $D_{t,t+1}(p_i)$ is the distance of p_i at t to the closest point on Γ_{t+1}, $\nabla\phi_t(x_{p_i}, y_{p_i})$ and $\nabla\phi_{t+1}(x_{p_i}, y_{p_i})$ are the gradients of p_i on ϕ_t and ϕ_{t+1} respectively. Equation (A.1) defines the search distance as the projection of the closest distance from p_i at t to the Γ_{t+1} onto the gradient direction of p_i on ϕ_t. A positive search distance corresponds to cell protrusion. Figure A.1 explains the search distance in Equation (A.1).

The tracking process can be achieved by computing search distances iteratively. The velocity v_i^t at marker p_i is defined as the distance between p_i on Γ_t and Γ_{t+1} on the gradient direction $\nabla\phi_t(p_i)$ computed from the tracking process and divided by the number of frames between time t and $t+1$ (which is one for my data). Positive velocity points to the direction of cell protrusion. The tracking algorithm is described in Algorithm 7. After boundary tracking, the velocities are extracted for each marker at different frames $\mathbf{v}_i = \{v_i^1, \ldots, v_i^{k-1}\}$ where k is the total number of frames.

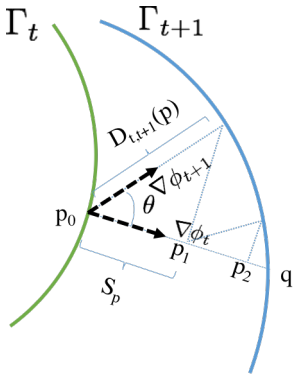


Figure A.1: Illustration of search distance S_{p_i} in Equation (A.1). p_0 is the location of a marker on Γ_t, while q is the corresponding location of the marker on Γ_{t+1}. The goal is to estimate the location of q. In this example, I use p_2 as the estimation the location q. Green curve and blue curve represent the cell boundaries Γ_t and Γ_{t+1} at time t and $t+1$ respectively; p is the sampled boundary point; $\nabla\phi_t$ and $\nabla\phi_{t+1}$ are the gradient directions of level sets ϕ_t and ϕ_{t+1} at point p respectively; S_p is the search distance at point p which is a projection of $D_{t,t+1}(p)$ onto the unit normal $\nabla\phi_t$; θ is the angle between $\nabla\phi_t(p_i)$ and $\nabla\phi_{t+1}(p_i)$ and $\cos\theta$ can be computed from the inner product of $\frac{\nabla\phi_{t+1}(p_i)}{|\nabla\phi_{t+1}(p_i)|}$ and $\frac{\nabla\phi_t(p_i)}{|\nabla\phi_t(p_i)|}$.

A.2 Principal Component Analysis (PCA)

Principal component analysis is a method which determines orthogonal directions which explain the maximum data variance. These directions are called principal components (PC)) (Wold et al., 1987; Abdi and Williams, 2010). PCA is different from dictionary learning as for dictionary learning, the dictionary atoms are usually not orthogonal with each other.

Consider the centered data points $x_i \in \mathbb{R}^p, i \in \{1, \ldots, n\}$, the projection of x_i onto a PC v is $v^T x_i$. The variance of the projected data is

$$\textbf{Var}(v^T X) = \frac{1}{n}\sum_{i}^{n}(v^T x_i)^2 = \frac{1}{n} v^T X X^T v, \tag{A.2}$$

where $X \in \mathbb{R}^{p\times n}$ is the data matrix, and each column represent a data point x_i. PCA can be formulated as,

$$\begin{aligned} \hat{v} = \underset{v}{\text{argmax}} \;\; & v^T X X^T v, \\ \text{s.t. } & v^T v = 1. \end{aligned} \tag{A.3}$$

Algorithm 7 Boundary Tracking

Input: Locations of sampled markers in 1st frame: $\{C_{p_i^1}\}$, $i \in \{1, ..., n\}$;
level set functions for different frames: $\phi_t,\ t \in \{1, \ldots, k-1\}$;
Output: velocities $\mathbf{v}_i = \{v_i^1, \ldots, v_i^k\}$;

let $v_i^0 = 0$;
for each frame t, $t \in \{1, \ldots, k-1\}$ **do**
  for each boundary point p_i^t **do**
    Compute gradient $\nabla\phi_t(p_i^t)$ and $\nabla\phi_{t+1}(p_i^t)$ for each p_i^t;
    let $\hat{p}_i^{t+1} = p_i^t$;
    let $\hat{v}_i^t = v_i^t$;
    while $\phi_{t+1}(\hat{p}_i^{t+1}) > \epsilon$ **do**
      $S_{p_i^t} = (D_{t,t+1}(\hat{p}_i^t)\frac{\nabla\phi_{t+1}(\hat{p}_i^t)}{|\nabla\phi_{t+1}(\hat{p}_i^t)|}) \cdot \frac{\nabla\phi_t(p_i^t)}{|\nabla\phi_t(p_i^t)|}$;
      $\hat{v}_i^t = \hat{v}_i^t + S_{p_i^t}$;
      $\hat{p}_i^{t+1} = \hat{p}_i^{t+1} + S_{p_i} \cdot \frac{\nabla\phi_t(p_i^t)}{|\nabla\phi_t(p_i^t)|}$;
    end while
    $v_i^t = \hat{v}_i^t$.
  end for
end for

The Lagrangian of Equation (A.3) is (Bellman, 1956),

$$L = v^T X X^T v + \lambda(I - v^T v). \tag{A.4}$$

Differentiating L respect to v and set to zero, we have,

$$(XX^T)v = \lambda v, \tag{A.5}$$

where v is an eigenvector of XX^T and the Lagrangian multiplier λ is the corresponding eigenvalue. The $1st$ PC v_1 is the eigenvector of the sample covariance matrix XX^T associated with the largest eigenvalue λ_1, similarly, the $2nd$ PC v_2 is eigenvector associated with the second largest eigenvalue λ_2 and so on.

For a given m, we have $V_m = [v_1 \ldots v_m] \in \mathbb{R}^{p \times m}$, then the original data X can be approximated as

$$\hat{X} = (V_m^+)^T V_m^T X = V_m^T (V_m^T X), \tag{A.6}$$

where $V_m^+ \in \mathbb{R}^{m \times p}$ is Moore-Penrose pseudoinverse (Albert, 1972). For orthogonal matrix V_m, $V_m^+ = V_m$ (Albert, 1972).

PCA can be used for dimensionality reduction. This has been widely used in computer vision applications such as face recognition (Turk and Pentland, 1991).

BIBLIOGRAPHY

Abdi, H. and Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433–459.

Aharon, M., Elad, M., and Bruckstein, A. (2006). K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322.

Albert, A. (1972). *Regression and the Moore-Penrose pseudoinverse*. Elsevier.

Aljabar, P., Heckemann, R. A., Hammers, A., Hajnal, J. V., and Rueckert, D. (2009). Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage*, 46(3):726–738.

Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202.

Bellman, R. (1956). Dynamic programming and lagrange multipliers. *Proceedings of the National Academy of Sciences of the United States of America*, 42(10):767.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300.

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 259–302.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

Boulanger, J., Kervrann, C., Bouthemy, P., Elbau, P., Sibarita, J., and Salamero, J. (2010). Patch-based nonlocal functional for fluorescence microscopy image sequences. *Medical Imaging, IEEE Transactions on*, 29(2):442–454.

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

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

Burridge, K. and Wennerberg, K. (2004). Rho and rac take center stage. *Cell*, 116(2):167–179.

Cao, T., Jojic, V., Modla, S., Powell, D., Czymmek, K., and Niethammer, M. (2013a). Robust multimodal dictionary learning. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pages 259–266. Springer Berlin Heidelberg.

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

Cao, T., Zach, C., Modla, S., Powell, D., Czymmek, K., and Niethammer, M. (2013b). Multi-modal registration for correlative microscopy using image analogies. *Medical image analysis*.

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

Chou, C.-R., Frederick, B., Mageras, G., Chang, S., and Pizer, S. (2013). 2d/3d image registration using regression learning. *Computer Vision and Image Understanding*, 117(9):1095–1106.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

Drori, I., Cohen-Or, D., and Yeshurun, H. (2003). Fragment-based image completion. volume 22, pages 303–312. ACM.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., et al. (2004). Least angle regression. *The Annals of statistics*, 32(2):407–499.

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

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

Fang, R., Chen, T., and Sanelli, P. C. (2013). Towards robust deconvolution of low-dose perfusion ct: Sparse perfusion deconvolution using online dictionary learning. *Medical image analysis*.

Fischl, B., Salat, D. H., van der Kouwe, A. J., Makris, N., Ségonne, F., Quinn, B. T., and Dale, A. M. (2004). Sequence-independent segmentation of magnetic resonance images. *Neuroimage*, 23:S69–S84.

Fisher, B., Perkins, S., Walker, A., and Wolfart, E. (1996). *Hypermedia image processing reference*. Wiley Chichester, UK.

Freeman, W. T., Jones, T. R., and Pasztor, E. C. (2002). Example-based super-resolution. *Computer Graphics and Applications, IEEE*, 22(2):56–65.

Friedman, J., Hastie, T., Höfling, H., Tibshirani, R., et al. (2007). Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1(2):302–332.

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

Guerrero, R., Pizarro, L., Wolz, R., and Rueckert, D. (2012). Landmark localisation in brain mr images using feature point descriptors based on 3d local self-similarities. In *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, pages 1535–1538. IEEE.

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

Hill, D. L., Batchelor, P. G., Holden, M., and Hawkes, D. J. (2001). Medical image registration. *Physics in medicine and biology*, 46(3):R1.

Horn, B. K. and Schunck, B. G. (1981). Determining optical flow. In *1981 Technical symposium east*, pages 319–331. International Society for Optics and Photonics.

Huang, J., Zhang, S., and Metaxas, D. (2011a). Efficient mr image reconstruction for compressed mr imaging. *Medical Image Analysis*, 15(5):670–679.

Huang, J., Zhang, T., and Metaxas, D. (2011b). Learning with structured sparsity. *The Journal of Machine Learning Research*, 12:3371–3412.

Huang, K. and Aviyente, S. (2007). Sparse representation for signal classification. *Advances in neural information processing systems*, 19:609.

Ibanez, L., Schroeder, W., Ng, L., and Cates, J. (2003). The ITK software guide.

Iglesias, J. E., Konukoglu, E., Zikic, D., Glocker, B., Van Leemput, K., and Fischl, B. (2013). Is synthesizing mri contrast useful for inter-modality analysis? In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pages 631–638. Springer.

Jaffe, A. B. and Hall, A. (2005). Rho GTPases: biochemistry and biology. *Annu. Rev. Cell Dev. Biol.*, 21:247–269.

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., and Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):881–892.

Kaynig, V., Fischer, B., Wepf, R., and Buhmann, J. (2007). Fully automatic registration of electron microscopy images with high and low resolution. *Microsc Microanal*, 13(Suppl 2):198–199.

Kim, M., Wu, G., Yap, P.-T., and Shen, D. (2012). A general fast registration framework by learning deformation–appearance correlation. *Image Processing, IEEE Transactions on*, 21(4):1823–1833.

Klein, S., Staring, M., Murphy, K., Viergever, M. A., Pluim, J. P., et al. (2010). Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging*, 29(1):196–205.

Kreutz-Delgado, K., Murray, J., Rao, B., Engan, K., Lee, T., and Sejnowski, T. (2003). Dictionary learning algorithms for sparse representation. *Neural computation*, 15(2):349–396.

Lemoine, D., Lussot, E., Legeard, D., and Barillot, C. (1994). Multimodal registration system for the fusion of mri, ct, meg, and 3d or stereotactic angiographic data. In *Medical Imaging 1994*, pages 46–56. International Society for Optics and Photonics.

Li, Y. and Osher, S. (2009). Coordinate descent optimization for $\ell 1$ minimization with application to compressed sensing; a greedy algorithm. *Inverse Probl. Imaging*, 3(3):487–503.

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

Machacek, M., Hodgson, L., Welch, C., Elliott, H., Pertz, O., Nalbant, P., Abell, A., Johnson, G. L., Hahn, K. M., and Danuser, G. (2009). Coordination of rho gtpase activities during cell protrusion. *Nature*, 461(7260):99–103.

Maintz, J. A. and Viergever, M. A. (1998). A survey of medical image registration. *Medical image analysis*, 2(1):1–36.

Mairal, J., Bach, F., Ponce, J., and Sapiro, G. (2009). Online dictionary learning for sparse coding. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 689–696. ACM.

Mairal, J., Bach, F., Ponce, J., Sapiro, G., and Zisserman, A. (2008). Supervised dictionary learning. In *NIPS*, pages 1033–1040.

Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., and Buckner, R. L. (2007). Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*, 19(9):1498–1507.

MATLAB (2012). *version 7.14.0 (R2012a)*. The MathWorks Inc., Natick, Massachusetts.

Miller, M. I., Trouvé, A., and Younes, L. (2006). Geodesic shooting for computational anatomy. *Journal of mathematical imaging and vision*, 24(2):209–228.

Mitchell, T. (1997). *Machine Learning*. McGraw Hill.

Modersitzki, J. (2009). *FAIR: flexible algorithms for image registration*, volume 6. SIAM.

Monaci, G., Jost, P., Vandergheynst, P., Mailhe, B., Lesage, S., and Gribonval, R. (2007). Learning multimodal dictionaries. *Image Processing, IEEE Transactions on*, 16(9):2272–2283.

Neal, R. M. and Hinton, G. E. (1998). A view of the em algorithm that justifies incremental, sparse, and other variants. In *Learning in graphical models*, pages 355–368. Springer.

Pennec, X., Cachier, P., and Ayache, N. (1999). Understanding the demons algorithm: 3d non-rigid registration by gradient descent. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI99*, pages 597–605. Springer.

Prince, J. L., Pham, D., and Tan, Q. (1995). Optimization of mr pulse sequences for bayesian image segmentation. *Medical Physics*, 22(10):1651–1656.

Qin, S. J. and Dunia, R. (2000). Determining the number of principal components for best reconstruction. *Journal of Process Control*, 10(2):245–250.

Redner, R. A. and Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM review*, 26(2):195–239.

Rice, J. (2006). *Mathematical statistics and data analysis*. Cengage Learning.

Ridley, A. J., Schwartz, M. A., Burridge, K., Firtel, R. A., Ginsberg, M. H., Borisy, G., Parsons, J. T., and Horwitz, A. R. (2003). Cell migration: integrating signals from front to back. *Science*, 302(5651):1704–1709.

Rohlfing, T. and Maurer, C. R. (2005). Multi-classifier framework for atlas-based image segmentation. *Pattern Recognition Letters*, 26(13):2070–2079.

Roy, S., Carass, A., and Prince, J. (2011). A compressed sensing approach for mr tissue contrast synthesis. In *Information Processing in Medical Imaging*, pages 371–383. Springer.

Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L., Leach, M. O., and Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: application to breast mr images. *Medical Imaging, IEEE Transactions on*, 18(8):712–721.

Scheffzek, K. and Ahmadian, M. R. (2005). Gtpase activating proteins: structural and functional insights 18 years after discovery. *Cellular and Molecular Life Sciences CMLS*, 62(24):3014–3038.

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual review of psychology*, 46(1):561–584.

Shi, Y., Wu, G., Song, Z., and Shen, D. (2012). Dense deformation reconstruction via sparse coding. In *Machine Learning in Medical Imaging*, pages 36–44. Springer.

Singh, N., Hinkle, J., Joshi, S., and Fletcher, P. T. (2013). A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction. In *ISBI*, pages 1219–1222.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43):15545–15550.

Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86.

van Tulder, G. and de Bruijne, M. (2015). Why does synthesized data improve multi-sequence classification? In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, pages 531–538. Springer.

Wachinger, C. and Navab, N. (2010). Manifold learning for multi-modal image registration. *11th British Machine Vision Conference (BMVC)*.

Wang, Q., Kim, M., Wu, G., and Shen, D. (2013). Joint learning of appearance and transformation for predicting brain mr image registration. In *Information Processing in Medical Imaging*, pages 499–510. Springer.

Wang, S., Zhang, D., Liang, Y., and Pan, Q. (2012). Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In *CVPR*, pages 2216–2223.

Wein, W., Brunke, S., Khamene, A., Callstrom, M. R., and Navab, N. (2008). Automatic ct-ultrasound registration for diagnostic imaging and image-guided intervention. *Medical image analysis*, 12(5):577.

Weiss, N. and Weiss, C. (2012). *Introductory statistics*. Addison-Wesley.

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

Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1):37–52.

Wright, J., Ma, Y., Mairal, J., Sapiro, G., Huang, T., and Yan, S. (2010). Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, 98(6):1031–1044.

Wright, J., Yang, A., Ganesh, A., Sastry, S., and Ma, Y. (2009). Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(2):210–227.

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

Yang, J., Wang, Z., Lin, Z., Shu, X., and Huang, T. (2012b). Bilevel sparse coding for coupled feature spaces. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2360–2367. IEEE.

Yang, J., Wright, J., Huang, T., and Ma, Y. (2010). Image super-resolution via sparse representation. *Image Processing, IEEE Transactions on*, 19(11):2861–2873.

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

Younes, L. (2010). *Shapes and Diffeomorphisms*, volume 171. Springer.

Younes, L., Arrate, F., and Miller, M. I. (2009). Evolutions equations in computational anatomy. *NeuroImage*, 45(1):S40–S50.

Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D. N., and Zhou, X. S. (2012a). Towards robust and effective shape modeling: Sparse shape composition. *Medical image analysis*, 16(1):265–277.

Zhang, S., Zhan, Y., and Metaxas, D. N. (2012b). Deformable segmentation via sparse representation and dictionary learning. *Medical Image Analysis*.

Zheng, Y., John, M., Liao, R., Boese, J., Kirschstein, U., Georgescu, B., Zhou, S. K., Kempfert, J., Walther, T., Brockmann, G., et al. (2010). Automatic aorta segmentation and valve landmark detection in c-arm ct: application to aortic valve implantation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, pages 476–483. Springer.

Zitova, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000.