Spring 5-16-2019

Deformable Image Registration for Hyperspectral Images

Rui Liao
Washington University in St. Louis

Washington University in St. Louis

McKelvey School of Engineering

Department of Electrical and Systems Engineering

Thesis Examination Committee:
Joseph A. O'Sullivan
Miaomiao Zhang
Jr-Shin Li

Deformable Image Registration for Hyperspectral Images

by

Rui Liao

A thesis presented to the McKelvey School of Engineering
of Washington University in partial fulfillment of the
requirements for the degree of

Master of Science

May 2019
Saint Louis, Missouri

Contents

List of Tables

List of Figures

Acknowledgments

I would first thank my research advisor, Prof. Joseph A. O'Sullivan for his valuable guidance and innumerable consultations I received. Dr.O always showed remarkable patience in answering my naive questions and guided me in a right direction of my thesis.

I also want to thank Prof. Miaomiao Zhang. I have learnt a lot from her Image Analysis course and also received valuable guidance from her related works.

Finally, I would like to thank Prof. Raymond E. Arvidson from the Earth and Planetary Science department, who inspired me for the thesis topic and gave me access to the resources I need.

Rui Liao

Washington University in Saint Louis
May 2019

ABSTRACT OF THE THESIS

Deformable Image Registration for Hyperspectral Images

by

Rui Liao

Master of Science in

Washington University in St. Louis, May 2019

Research Advisor: Professor Joseph A. O'Sullivan

Image registration is one of the basic image processing operations in remote sensing. A hyperspectral image has two spatial dimensions and one spectral dimension. There are many hyperspectral sensors used in remote sensing. Traditional intensity-based registration methods may fail for hyperspectral images because of the different spectral sensitivities for different sensors. In addition, not all spectral bands are required to achieve accurate registration. This thesis develops a modification of the large deformation diffeomorphic metric mappings (LDDMM) algorithm in order to deal with the challenges when applied to hyperspectral images. The transformation generated by our method that deforms one image to match the other is differentiable, isomorphic and invertible. We also propose a mutual information based band selection algorithm to reduce the data redundancy of the hyperspectral images. The approach is applied to two hyperspectral images from OMEGA instrument, with a better matching result than original LDDMM method with respect to mutual information.

Chapter 1

Introduction

Image registration is the process of aligning two or more images of the same scene taken at different times, from different viewpoints or by different sensors [21]. For the two images registration problem, we will denote one image as a fixed image to provide a reference to the other called a moving image. The registration algorithm will generate an optimal transformation so that corresponding coordinate points in two images will be matched. It is a classical problem in several image processing applications in which the final information is obtained from the combination of various data sources such as motion detection, image fusion, and multichannel image restoration. Registration methods are very sensitive to different image conditions. Given the diversity of the data, it is unlikely that a single registration scheme will work satisfactorily for all different applications.

The purpose of this thesis is to provide an automatic image registration algorithm for hyperspectral images obtained from OMEGA [4]. In this section, we will first introduce some existing methods for registration problems, then focus on related topics for hyperspectral data.

1.1 Image Registration Methodology

Image registration is widely used in remote sensing, medical imaging, computer vision, and other fields. In general, its applications can be divided into two main groups according to different matching focuses: Intensity-Based and Feature-based.

1.1.1 Intensity-Based Methods

Intensity-based methods, sometimes called correlation-like methods or template matching [11] align images without attempting to detect salient objects. A classical method like cross-correlation (CC) is used for matching directly image intensities, without any structural analysis. Consequently, is sensitive to the intensity changes, introduced by noise, different illumination conditions or by using different sensor types.

These methods share the same framework that minimizes the distance between the moving and fixed images. The core part for different intensity-based methods is the definition of distance or similarity of images. Here are three commonly used distance functions.

1. *Correlation-like methods.* The similarity of fixed and moving images is defined as

$$CC(i,j) = \frac{\sum(I - E[I])(J_{i,j} - E[J_{i,j}])}{\sqrt{\sum(I - E[I])^2 \sqrt{\sum(I_{i,j} - E[I_{i,j}])^2}}}, \tag{1.1}$$

   where i, j is the pixel coordinates for image I and J.

   This measure of similarity is computed for window pairs from the fixed and moving images and its maximum is searched. Two main drawbacks of the correlation-like methods are the flatness of the similarity measure maxima (due to the self-similarity of the images) and high computational complexity. Most of the methods deal with translation-only images, more complicated geometric deformations will dramatically increase the computational load.

2. *Cross-power Spectrum.* This registration method is performed in the Fourier domain. It computes the cross-power spectrum of the fixed and moving images and looks for the location of the peak in its inverse [5].

$$\frac{\mathcal{F}(I)\mathcal{F}(J)^*}{|\mathcal{F}(I)\mathcal{F}(J)^*|} = e^{2\pi j(ux_0 + vy_0)}, \tag{1.2}$$

   where $\mathcal{F}$ represents the 2D Fourier transformation and $*$ is the conjugate function.

   This method is strongly appreciated when an acceleration of the computational speed is needed or the images are acquired under varying conditions or they are corrupted

by frequency-dependent noise. The method shows strong robustness against the correlated and frequency dependent noise and non-uniform, time-varying illumination disturbances.

3. *Mutual Information.* Mutual information is a concept from information theory, which measures the dependence between two random variables. The definition is given by

$$MI(I_0; I_1) = H(p(I_0)) + H(p(I_1)) - H\left(p(I_0), p(I_1)\right), \tag{1.3}$$

where $p(I_0), p(I_1)$ is the marginal distribution (histogram) of image I_0 and I_1 and H represent entropy. This method represents the leading technique in multimodal registration and we will discuss it in detail in Section 2.1.

1.1.2 Feature-Based Methods

The majority of the feature based methods consists of the following four steps [21]:

1. *Feature detection.* Salient and distinctive objects (closed-boundary regions, edges, contours, line intersections, corners, etc.) are manually or, preferably, automatically detected. For further processing, these features can be represented by their point representatives (centers of gravity, line endings, distinctive points), which are called control points (CPs) in the literature.

2. *Feature matching.* In this step, the correspondence between the features detected in the moving image and those detected in the fixed image is established. Various feature descriptors and similarity measures along with spatial relationships among the features are used for that purpose.

3. *Transformation estimation.* The type and parameters of the so-called mapping functions or transformations, aligning the moving image with the fixed image, are estimated. The parameters of the mapping functions are computed by means of the established feature correspondence.

4. *Image deformation/transformation.* The moving image is deformed by means of the mapping functions. Image intensities in non-integer coordinates are computed by the appropriate interpolation technique.

The implementation of each registration step has its typical problems. For the detection step, the required features should be distinctive objects, which are frequently spread over images and must have enough common elements in both fixed and moving images. In the feature matching step, a fatal problem may be caused by image degradation. Physically corresponding features can be dissimilar due to the different imaging conditions and/or due to the different spectral sensitivity of the sensors.

Automatic feature detection and matching can be achieved perfectly by machine learning [16] [9]. However, a sufficiently clean training data set is required and may also involve a lot of manual labeling. But the concept of using feature information in images for registration is still considered as an accurate and fast approach to solving the problem.

1.2 Transformation Model

The transformation model is chosen to properly align two images, it describes how pixel locations move after deformation. There are typically two types of transformation model used in registration problems: parameterized and non-parameterized.

1.2.1 Parameterized Transformation

For a parameterized transformation model, the function can be determined by some low dimensional coefficients and applied on the entire image. One of the most frequently used models is affine transformation, which consists of rotation, translation, shear, and scaling. Figure 1.1 shows the result of applying the affine transformation. The advantages of using these models are they can be easily applied and represented. After setting up the metric between images, the model can be solved by parameter estimation or search to achieve the minimum distance. However, these models treat the whole image scene equally, and apply

the same transformation on every pixel. But sometimes in medical or remote sensing images, deformation only happens on some local area, for example, in a CT scan of the leg, the bone should not share the same transformation with the muscle tissue around it. For the OMEGA data we are dealing with, the image is kind of slim (2288×32), small rotation from the bottom will cause a large empty area at the top. In these cases, we can not simply use low dimensional parameterized transformation models.

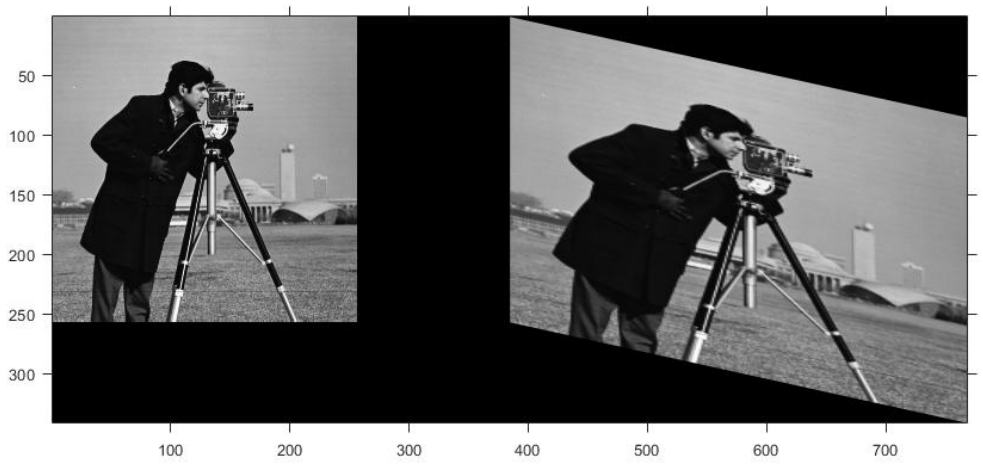

Figure 1.1: Affine Transformation

1.2.2 Non-Parameterized Transformation

Another approach to the registration of images with considerable complexity and/or local distortions is not to use any parametric mapping functions, where the estimation of the geometric deformation is reduced to the search for the best parameters.

The images are viewed as pieces of a rubber sheet, on which external forces stretching the image and internal forces defined by stiffness or smoothness constraints are applied to bring them into alignment with the minimal amount of bending and stretching [2]. This idea is often called *elastic registration*. The external forces in implementation are often considered as a vector field with the same spatial dimension with the moving image. For each pixel, a two-dimensional vector will represent the direction and value of the external force, an

example is shown in Figure 1.2. For each coordinates in image domain, a 2-dimensional vector represents the direction and distance this pixel will move.

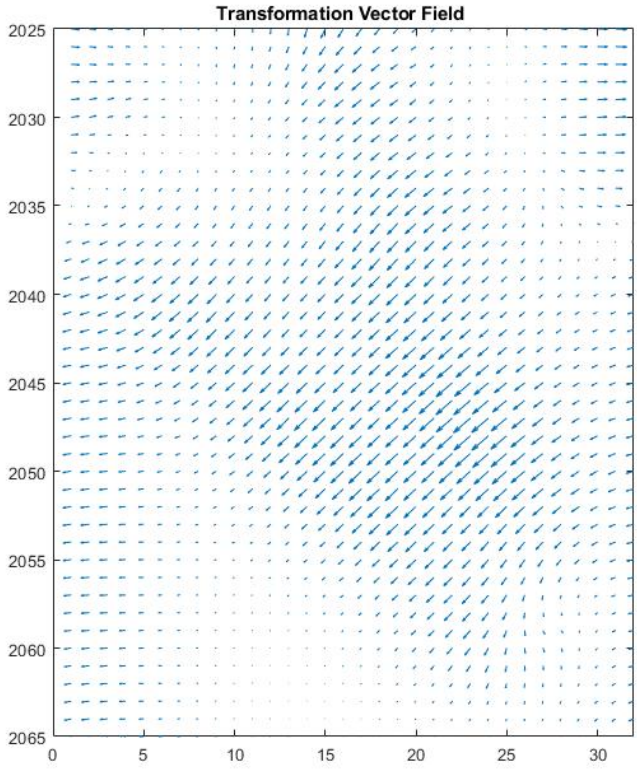

Figure 1.2: Vector Field on Image Domain

However, when using this transformation model, the regularization term or smoothness constraint will play an important part. Without any constraint, the external forces can move any pixel to any place, which may destroy the topology of the original image. For example, we do not want the registration algorithm to move the meteor crater initially on the left and align it to another crater on the right. Also, the complexity of this model will bring a lot of computational cost into the algorithm.

This idea can be further extended to diffeomorphisms, that require the transformation to be differentiable and invertible. In this case, a path or a curve is defined to represent how a transformation moves with time on the manifold it lives in. In addition, with this curve, a sequence of images to show the process of deformation is available for view.

1.3 Hyperspectral Imagery

Hyperspectral data can be regarded as a three-dimensional image cube in which the first two dimensions are spatial and the third dimension is spectral. Figure 1.3 illustrates the structure of the hyperspectral data set OMEGA. Hyperspectral images are important for remote sensing applications such as mineralogy, geology, agriculture, and surveillance sensing.

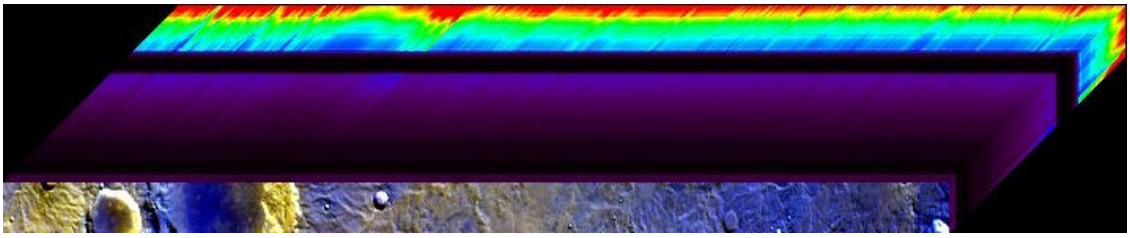

Figure 1.3: Visualization of OMEGA Data in SWIR Channel

OMEGA (Observatoire pour la Minralogie, lEau, les Glaces et lActivit) is a visible and near-IR mapping spectrometer, operating in the spectral range 0.38 - 5.1 μm. Combining imaging and spectrometry, it will study the mineralogical and molecular composition of the surface and atmosphere of Mars through the spectral analysis of the diffused solar light and surface thermal emission. OMEGA will provide global coverage at medium-resolution (2-5 km) for altitudes from 1500 km to 4000 km, and high-resolution (< 350 m) spectral images of selected areas, amounting to a few percents of the surface, when observed from near-periapsis (< 300 km altitude)[4].

These two optical units are supposed to be co-aligned. However, after years of operation on the orbit of Mars, images from the two channels become deformed and some areas are mismatched. Our purpose in this thesis is to align these two image cubes from OMEGA as a pre-processing of the data that can be extended to further applications.

1.4 Thesis Layout

In Chapter 2, we discuss a mutual information-based band selection method we use to reduce the redundancy of the OMEGA data and choose a good pair of bands for registration. Results

and comparison with other methods, such as PCA and entropy-based band selection will be included.

In Chapter 3, the original LDDMM algorithm is introduced and along with the mathematical foundation of this methodology. In the later part of this chapter, we will present our modified LDDMM algorithm to enhance the performance of the original LDDMM method. This thesis propose an augmented energy function with an additional penalty term for matching the edge maps of the given moving and fixed images. This modification significantly reduces the number of local minimums for the original energy function that will give a more robust performance to the gradient descent algorithm (we are using gradient descent algorithm to minimize the energy function). Results and comparison are presented in the last section.

Chapter 2

Band Selection for Hyperspectral Image Registration

Hyperspectral imagery generally consists of enormous amounts of spectra or bands which measure intensity at various wavelength. For the OMEGA data we are using, its optical part contains two co-aligned units: a Visible Channel (VNIR) analyses the light from 0.36 to 1.07 μm; a Near Infrared Channel, named SWIR (Short Wavelength IR Channel) disperses the light through two spectrometers from 0.93 to 2.7 μm and 2.6 to 5.2 μm [4]. For VNIR channel, it generates an image cube with dimension $2288 \times 32 \times 103$, where the third dimension represents numbers of bands. And for SWIR, the data dimension is $2288 \times 32 \times 113$. Registration through all pairs of these bands is known to be computationally hard. In this part, we are using a mutual information based band selection method to reduce data redundancy of OMEGA data.

This part is optional because of the robustness of our registration method. An arbitrary pair of bands could give similar result with the pair of band chose by our band selection algorithm. Here we provide one choice of data reduction for OMEGA data set.

2.1 Basic Math Review for Information Theory

In Shannon's information theory, entropy measures information content in terms of uncertainty and mutual information is used to quantify the degree of independence between two random variables. These two concepts provide a methodology to reduce data redundancy

and useless information. Let us introduce some information theory concepts and properties [8] in the following parts.

2.1.1 Information Entropy

The entropy of a random variable X is expressed as

$$H(X) = -\sum_{x \in \Omega} p(x) \log p(x), \tag{2.1}$$

where $p(x)$ represents the mass probability of an event $x \in \Omega$ from a finite set of possible values. The entropy of a random variable is a measure of the amount of information required on the average to describe a random variable.

Some papers [1] ([17],chapter 6) directly use the information entropy as a criterion to select bands in hyperspectral images. The image cube is sorted in descending order subjects to the entropy $H(I(\lambda_i))$ of each individual band or wavelength interval λ_i.

$$H(\lambda_i) = -\sum_{k=1}^{m} p(\phi_k^i) \log p(\phi_k^i), \tag{2.2}$$

$$p(\phi_k^i) \doteq p(min_k^i \leq I(\lambda_i) < max_k^i), \tag{2.3}$$

where the probability mass function $p(\phi_k^i)$ is estimated via a histogram where each bin k used to estimate a probability for a range of values of $I(\lambda_i)$ is defined by $\{min_k^i, max_k^i\}$. m is the number of bins used in each histogram [13]. Band selection is performed by choosing the bands with higher entropy values.

However, this information measurement does not always, as often assumed, coincide with the requirement of image registration. For image registration between two hyperspectral cubes, a choice is to select one particular band for each cube and register them in order to reduce computational complexity. Thus, we are more interested in how similar are these two selected bands.

From Equation 2.2, it is shown that the entropy is a function of a single image cube $I(\lambda_i)$. In other words, $H(I(\lambda_i))$ is calculated on a single signal, with no reference to another objective. This means that the amount of information measured by the entropy lacks a point of reference or benchmark [14]. Hence, there is no guarantee that the amount of entropy always matches the information content useful for registration.

2.1.2 Mutual Information

To improve the entropy-based band selection methods, a better approach is to introduce the concept of mutual information (MI). The mutual information of two random variables is a measure of the mutual dependence between the two variables. More specifically, it quantifies the "amount of information" obtained about one variable by observing the other random variable. The definition of MI between X, Y is

$$I(X;Y) = \sum_{x,y\in\Omega} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \tag{2.4}$$

$$= D(p(x,y)||p(x)p(y)), \tag{2.5}$$

where $p(x,y)$ is the joint probability distribution of X, Y and $p(x), p(y)$ are their marginal distribution, D is the Kullback-Leibler distance.

From Equation 2.4 and 2.1, it is easy to find out the relationship between the mutual information and the entropy

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{2.6}$$

$$= H(X) - H(X|Y) \tag{2.7}$$

$$= H(Y) - H(Y|X), \tag{2.8}$$

where $H(X,Y)$ is the joint entropy and $H(X|Y), H(Y|X)$ is the conditional entropy. Definitions given by Shannon are shown as

$$H(X,Y) = -\sum_{x,y} p(x,y) \log p(x,y) \tag{2.9}$$

$$H(X|Y) = -\sum_{x,y} p(x,y) \log p(x|y). \tag{2.10}$$

Mutual information is always non-negative for two random variables, being zero when the variables are statistically independent. A higher value of MI indicates a higher dependency between the variables. However, it is still not a wise approach to calculate all pairs of bands between two image cubes and choose the pair with the largest value. For example, for OMEGA data, computations for 103×113 pairs of images are needed.

2.2 Hyperspectral Band Selection through Information-theoretic Analysis

Based on the discussion of entropy and mutual information in section 2.1, this paper presents a mutual information based band selection algorithm that can choose a good pair of bands from two image cubes for registration within few seconds.

2.2.1 MI-based Band Selection with Pre-processing

The goal for band selection is to find a good pair of bands with mutual information closest to the maximum value.

The preprocessing part of our method is the same as the entropy-based method, that is sorting all bands in descending order with respect to the entropy. Instead of choosing the band with the largest value of entropy, we store a small percentage of bands with higher values, denotes as a reduced-dimension cube, for example, 10%. Then, for these two reduced-dimension cubes, we calculate the mutual information for each pair of bands and select the pair with the largest value.

The detail of this algorithm is shown in the following Algorithm 1.

Algorithm 1 MI-based band selection algorithm.

Require: The moving image cube, M; The fixed image cube, F; Numbers of bands stored for each image cube after preprocessing, N.

Ensure: Extract one band for each cube that their mutual information is the largest among stored bands, $\hat{M}_N, \hat{F}_N$; The value of the largest mutual information, $MI(\hat{M}_N, \hat{F}_N)$;

1: Calculate the entropy for each band in M and F;
2: Sort each cube M, F by entropy and store only top N bands into matrices M_N, F_N;
3: Construct similarity matrix $S_{N\times N}$, for each entry (i, j), $S(i, j) = MI(M_N(i), F_N(j))$;
4: Find the index of the pair of bands with the largest value in S, i^*, j^* and its value, MI^*;
5: Return the pair of bands, $\hat{M}_N = M_N(i^*)$, $\hat{F}_N = F_N(j^*)$ and $MI(\hat{M}_N, \hat{F}_N) = MI^*$;

Proof. It is trivial for the initial condition when $N = 1$, that is we are choosing the band with the largest value of entropy for each image cube. Because there are only two bands, it is not necessary to compute the mutual information between them.

Suppose for $N = m$, after running the algorithm, it will return the optimal pair of band $\hat{M}_m$, $\hat{F}_m$ and their mutual information MI^*_m. According to step 3 in Algorithm 1, the similarity matrix should be $S_{m\times m}$ and $MI^*_m = \max(S_{m\times m})$. When $N = m+1$, we have that

$$S_{m\times m} \subset S_{(m+1)\times(m+1)}, \tag{2.11}$$

$$MI^*_m \leq MI^*_{m+1}. \tag{2.12}$$

Therefore, according to Equation 2.12, mutual information is non-decreasing as the number of bands N increases. When N grows to the maximum number of bands, the algorithm just performs like calculating the mutual information for each pair of bands and return the true optimal maximum value. □

2.2.2 Experimental Result

In section 2.2.1, we presented our method in detail and showed proof of correctness. The experimental results will consist of comparing our method with the entropy-based method using a OMEGA dataset.

First, we test the entropy-based method with an OMEGA data, the dimension of the VNIR channel is $2288 \times 32 \times 103$ and $2288 \times 32 \times 113$ for the SWIR channel. The whole image is too large to display in this thesis, so we only present two areas of the whole scene with some easily recognized features in Figure 2.1.

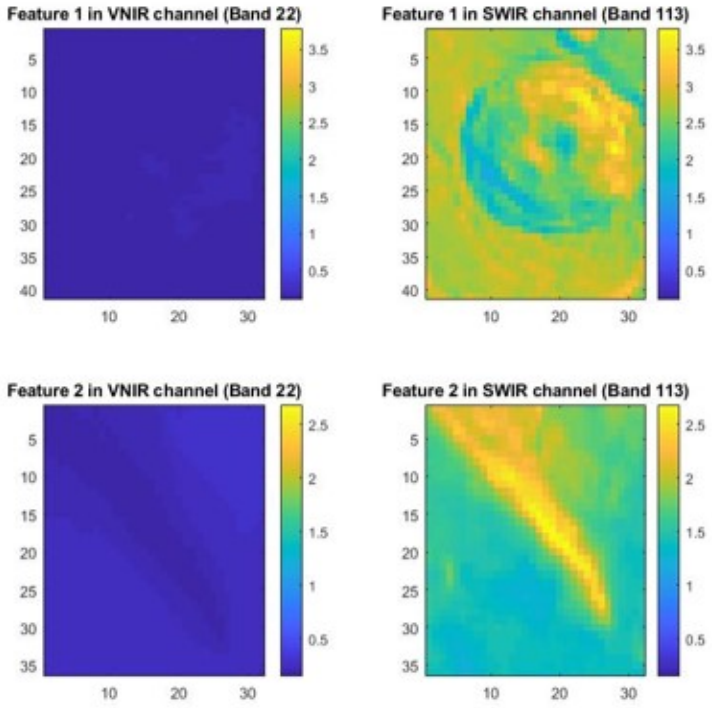

Figure 2.1: Entropy-based band selection result

In Figure 2.1 the left two figures are from the VNIR channel selected by the entropy-based method. Both corresponding images from two channels illustrate the same feature, but the intensity values vary a lot with the other. The mean intensity value for the band in VNIR channel is around 0.25, but for SWIR channel, this value is about 0.8. In terms of mutual information, the value is 0.015, which is not at an ideal level for image registration. Results from our MI-based band selection method is shown in Figure 2.2. The mutual information between these two bands is 1.73, which is about 114 times larger than the value we get for the entropy-based method.

Figure 2.3 is a plot showing that mutual information is a non-decreasing function of the number of bands we stored. When the number of bands reaches 45 (around 44% of all bands), the MI we get converges to the global optimal MI value. From experiments, an

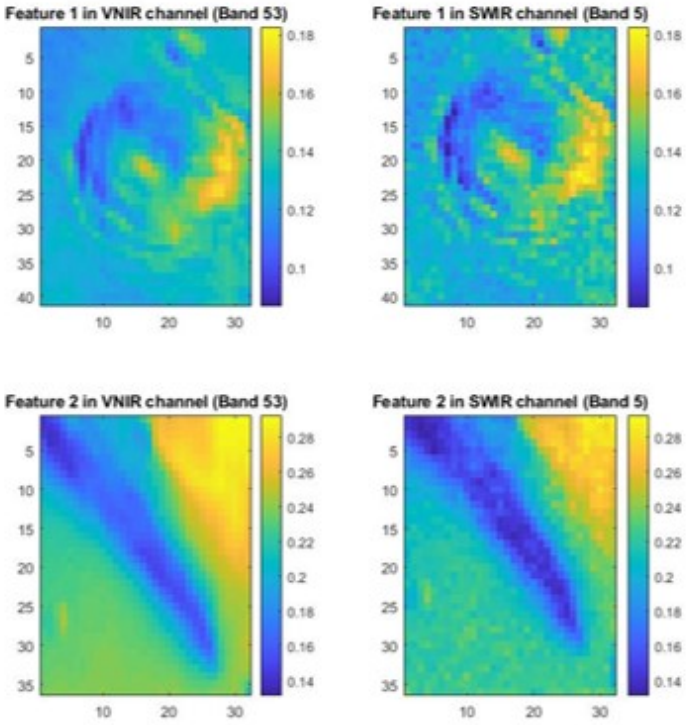

Figure 2.2: MI-based band selection result

approximate choice for percent of bands stored is 50%. We also compare our band selection result with other dimension reduction method, such as principal component analysis (PCA). The red dash line in Figure 2.3 shows the value of mutual information for images after PCA, which is only 0.5067. The reason for this low mutual information acquired by PCA is that it reduces the dimension only based on the distribution of one data set without considering the other.

Figure 2.4 shows the function of computation time for the different number of bands. From the previous discussion, a choice for the percentage of bands stored is 50%, and the corresponding computation time is about 2 seconds. Compare with calculating all pairs of bands in both cubes which takes 9 seconds to finish, our MI-based method only cost 22.2% of the time but returns exactly the same optimal result.

In conclusion, our mutual information based band selection algorithm provides a suitable approach that requires limited computation time to choose two corresponding bands from

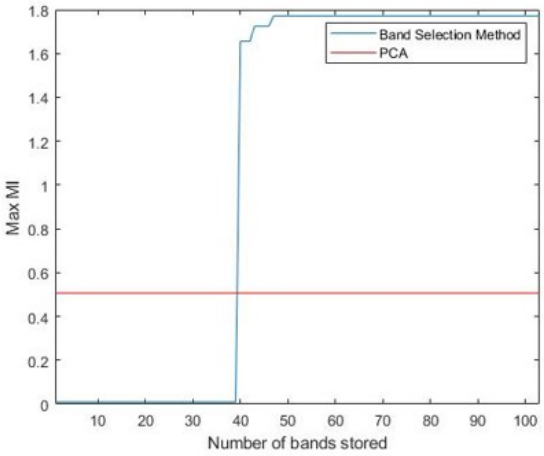

Figure 2.3: Correctness of MI-based method

each image cube in OMEGA dataset. For the next section, we will use the results generated in this part for the discussion of image registration problem.

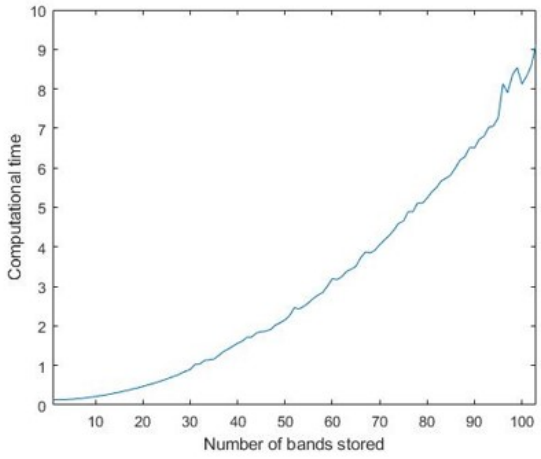

Figure 2.4: Computational Cost for Selecting Different Number of Bands

Chapter 3

Deformable Image Registration for Hyperspectral Image

In this chapter, we first discuss the large deformation diffeomorphic metric mappings (LDDMM) method that is widely applied in the medical image area [3]. Diffeomorphic image registration provides smooth and invertible smooth spatial correspondences between images. However, in the original LDDMM method, the sum of squared differences is used for evaluating the performance of the registration result, which is not a suitable criterion for hyperspectral imagery. In the second section of this chapter, we present our modified LDDMM method.

3.1 LDDMM

Both mathematical proof and experiment results are presented in this section.

3.1.1 Math Background and Notation

There are a few fundamental ingredients in LDDMM and our goal is to review them in this section without including the full mathematical proof. That is we will only introduce the notions that are needed for the understanding of the rest of this thesis.

1. A *diffeomorphism* of Ω is an invertible transformation from Ω to itself, which is continuously differentiable (C^1) with a differentiable inverse. The identity map, id, is a diffeomorphism and the inverse of any of them is a diffeomorphism as well. For any diffeomorphisms φ and any object $m \in \Omega$, $\varphi \cdot (\psi \cdot m) = (\varphi \circ \psi) \cdot m$. Properties above imply that the set of diffeomorphisms of Ω, denoted Diff(Ω) has a *group structure*.

2. *Displacement vector field* u is defined as $u : \Omega \to \mathbb{R}^d$ such that $\varphi(x) = x + u(x)$ for all points $x \in \Omega$. The transformation template image I_0 applied with this displacement field then becomes $I_0(\varphi(x)) = I_0(x + u(x))$. This transformation model is an early attempt to compute high-dimensional non-rigid transformations resulted in the development of the elastic matching strategy [6]. In order to meet our diffeomorphism property, we need to introduce the concept of a velocity field to make the transformation differentiable.

3. *Time-dependent vector fields.* A time-dependent vector field on Ω is a function:

$$v : [0, 1] \times \Omega \to \mathbb{R}^d \tag{3.1}$$

$$(t, x) \mapsto v(t, x). \tag{3.2}$$

   We will also use $v_t(x)$ to denote $v(t, x)$. The time interval is scale to $[0, 1]$. The corresponding ordinary differential equation (ODE) is

$$\frac{dy}{dt} = v_t(y). \tag{3.3}$$

   The flow of this ODE generates the evolution of a curve ϕ^v which depends on time and space

$$\frac{\partial \phi^v}{\partial t}(t, x) = v_t(\phi^v_t(x)). \tag{3.4}$$

   The initial point of this curve at time $t = 0$ is the identity transformation $id(x) = x, \ \forall x \in \Omega$. The endpoint at time $t = 1$ is the diffeomorphism or displacement field that has been discussed before $\varphi(x) = \phi(1, x), \ \forall x \in \Omega$. We also use the notation ϕ^v_t for $\phi(t, x)$ to show the dependence of ϕ on v. In this way, a displacement field or a diffeomorphism φ can be generated by integral all the time-dependent vector field v_t via Equation 3.4 from time $t = 0$ to $t = 1$.

Let the notation $\phi_{s,t} : \Omega \to \Omega$ denote the composition $\phi_{s,t} = \phi_t \circ (\phi_s)^{-1}$. The interpretation of $\phi_{s,t}(y)$ is that it is the position at time t of an object that is at position y at time s. In this case, the final diffeomorphism $\varphi = \phi_{1,0}$, and the initial transformation is $id = \phi_{0,0}$.

4. *Dot product over vector field.* We now introduce a dot product norm over vector fields. More precisely, we will consider a space V of vector fields on Ω and a dot product norm $||v||_V^2 = \langle v, v \rangle_V$.

   The norm on V is defined through a differential operator L (denoting the adjoint operator as $L^\dagger$) given by

$$||v||_V^2 = \int_\Omega (Lv)^T (Lv) dx, \tag{3.5}$$

$$\langle v, v \rangle_V \doteq (Lv, Lv) = (L^\dagger Lv, v), \tag{3.6}$$

   where $(,)$ is the usual L^2-product for square integrable vector fields on Ω. In our implementation, however, the inverse of the operator $K = (L^\dagger L)^{-1}$ is used most of the time. This compact self-adjoint operator K is defined by

$$(v, w) = \langle Kv, w \rangle_V. \tag{3.7}$$

   Together with Equation 3.6, for any smooth vector field $v \in V$, we have

$$K(L^\dagger L)v = v. \tag{3.8}$$

   This definition of norm in vector space V enforces a smoothness constraint that guarantees V as a smooth, compactly-supported vector field. The technical details associated with this construction of K and L are discussed in [18], [10]. And later in this thesis, we will present the numerical implementation for the operator K.

3.1.2 Variational Minimization of the Energy Function

The image registration problem can be expressed as an optimization problem defined as

$$\hat{v} = \underset{v:\dot{\phi}_t = v_t(\phi_t)}{\arg\min} \left(\underbrace{\int_0^1 ||v_t||_V^2 dt}_{E_1(v)} + \underbrace{\frac{1}{\sigma^2} ||I_0 \circ \phi_{1,0}^v - I_1||_{L^2}^2}_{E_2(v)} \right). \tag{3.9}$$

The so-called energy function is the term inside the argmin function in Equation 3.9, which consists of two parts. The second term $E_2(v)$ evaluates the metric between the deformed moving image and the fixed image using standard L^2 norm defined on Ω. The first term $E_1(v)$ is a measurement of smoothness defined on the vector space V.

The optimizer $\hat{v}$ of this energy function can generate the optimal transformation or diffeomorphism $\varphi = \phi_1^v$ by integration via Equation 3.4 with initial condition $\phi_0 = id$. Also, the optimal solution guarantees that the flow of transformation $\phi_{s,t}^v$ satisfies a sufficient amount of smoothness on the elements of the vector space V. To solve this variational optimization problem, we need to compute the variation of the transformation $\phi_{1,0}^v$ first.

Theorem 1. *The variation of transformation $\phi_{s,t}^v$ when $v \in L^2([0,1], V)$ is perturbed along $h \in L^2([0,1], V)$ for a small amount is given by:*

$$\partial_h \phi_{s,t}^v = D\phi_{s,t}^v \int_s^t (D\phi_{s,u}^v)^{-1} h_u \circ \phi_{s,u}^v du, \tag{3.10}$$

where $D\phi_{s,t}^v$ is the Jacobian of the transformation $\phi_{s,t}^v$.

Proof. From calculus of variation, we first have

$$\partial_h \phi_{s,t}^v = \lim_{\epsilon \to 0} \frac{\phi_{s,t}^{v+\epsilon h} - \phi_{s,t}^v}{\epsilon}. \tag{3.11}$$

According to Equation 3.4, the derivative of the perturbed transformation $\phi_{s,t}^{v+\epsilon h}$ at time t is

$$\frac{d}{dt}\phi_{s,t}^{v+\epsilon h} = v_t \circ \phi_{s,t}^{v+\epsilon h} + \epsilon h_t \circ \phi_{s,t}^v. \tag{3.12}$$

Substitute Equation 3.12 and 3.4 into Equation 3.11, an ordinary differential equation (ODE) can be derived as

$$\frac{d}{dt}\partial_h \phi^v_{s,t} = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \cdot \left[\frac{d}{dt}\phi^{v+\epsilon h}_{s,t} - \frac{d}{dt}\phi^v_{s,t} \right] \tag{3.13}$$

$$= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[v_t \circ \phi^{v+\epsilon h}_{s,t} + \epsilon h_t \circ \phi^{v+\epsilon h}_{s,t} - v_t \circ \phi^v_{s,t} \right] \tag{3.14}$$

$$= \lim_{\epsilon \to 0} \frac{\partial}{\partial \epsilon} \left[v_t \circ \phi^{v+\epsilon h}_{s,t} \right] + h_t \circ \phi^{v+\epsilon h}_{s,t} \tag{3.15}$$

$$= D_{\phi^v_{s,t}} v_t \cdot \partial \phi^v_{s,t} + h_t \circ \phi^{v+\epsilon h}_{s,t}. \tag{3.16}$$

By solving the ODE 3.16, we could show that the expression in Equation 3.10 is the correct solution with condition $\partial_h \phi^v_{s,s} = 0$. This is also trivial to directly compute this variation of $\phi^v_{s,t}$ by using the fact that

$$\frac{d}{dt} D\phi^v_{s,t} = D_{\phi^v_{s,t}} v_t D\phi^v_{s,t}, \tag{3.17}$$

which simply comes from the space differential of $\frac{d}{dt}\phi^v_{s,t} = v_t \circ \phi_{s,t}$. □

Theorem 2. *Given a continuously differentiable idealized fixed image I_0 and a moving image I_1, then $\hat{v} \in L^2([0,1], V)$ for inexact matching of I_0 and I_1 is formulated as*

$$\hat{v} = \underset{v \in L^2([0,1],V)}{\operatorname{arginf}} \; E(v) \doteq \left(\underbrace{\int_0^1 ||v_t||^2_V dt}_{E_1(v)} + \underbrace{\frac{1}{\sigma^2} ||I_0 \circ \phi^v_{1,0} - I_1||^2_{L^2}}_{E_2(v)} \right). \tag{3.18}$$

The Euler-Lagrange equation for the matching energy function at time t is given by

$$2\hat{v_t} - K\left(\frac{2}{\sigma^2} \left| D\phi^{\hat{v}}_{t,1} \right| \nabla J^0_t (J^0_t - J^1_t) \right) = 0, \tag{3.19}$$

where $J^0_t \doteq I_0 \circ \phi_{t,0}$, $J^1_t \doteq I_1 \circ \phi_{t,1}$.

Proof. Let the velocity $v \in L^2([0,1], V)$ be perturbed by an ϵ amount direction $h \in L^2([0,1], V)$. The variation $\partial_h E(v)$ of the energy function is defined by

$$\partial_h E(v) = \lim_{\epsilon \to 0} \frac{\partial E(v + \epsilon h)}{\partial h}. \tag{3.20}$$

The variation of $E_1(v)$ is given by

$$\partial_h E_1(v) = \lim_{\epsilon \to 0} \frac{\partial}{\partial h} \int_0^1 \langle v_t + \epsilon h_t, v_t + \epsilon h_t \rangle_V^2 dt \tag{3.21}$$

$$= 4 \int_0^1 \langle v_t, h_t \rangle_V dt. \tag{3.22}$$

The variation of $E_2(v)$ is given by

$$\partial_h E_2(v) = \frac{2}{\sigma^2} \langle I_0 \circ \phi_{1,0}^v - I_1, DI_0 \circ \phi_{1,0}^v \partial_h \phi_{1,0}^v \rangle_{L^2}. \tag{3.23}$$

The term $\partial_h \phi_{1,0}^v$ is the variation of the transformation which is given by Theorem 1. Substitute Equation 3.10 into Equation 3.23, we then have

$$\partial_h E_2(v) = \frac{2}{\sigma^2} \langle I_0 \circ \phi_{1,0}^v - I_1, DI_0 \circ \phi_{1,0}^v \cdot (-D\phi_{1,0}^v \int_0^1 (D\phi_{1,t}^v)^{-1} h_t \circ \phi_{1,t}^v dt) \rangle_{L^2}. \tag{3.24}$$

By collecting $D(I_0 \circ \phi_{1,0}^v) = DI_0 \circ \phi_{1,0}^v D\phi_{1,0}^v$, Equation 3.24 becomes

$$\partial_h E_2(v) = \frac{2}{\sigma^2} \langle I_0 \circ \phi_{1,0}^v - I_1, D(I_0 \circ \phi_{1,0}^v) \cdot (D\phi_{1,0}^v)^{-1} \cdot (-D\phi_{1,0}^v \int_0^1 (D\phi_{1,t}^v)^{-1} h_t \circ \phi_{1,t}^v dt) \rangle_{L^2} \tag{3.25}$$

$$= \frac{-2}{\sigma^2} \int_0^1 \langle I_0 \circ \phi_{1,0}^v - I_1, D(I_0 \circ \phi_{1,0}^v)(D\phi_{1,t}^v)^{-1} h_t \circ \phi_{1,t}^v \rangle_{L^2} dt. \tag{3.26}$$

Setting $\phi_{1,t}^v(y) = x$, then $\phi_{t,1}(x) = y$, which gives the Jacobian change of variables $|D\phi_{t,1}^v| dx = dy$. Thus, we have $\phi_{1,0}^v \circ \phi_{t,1}^v = \phi_{t,0}^v$ and substituting in above equation, we will have:

$$\partial_h E_2(v) = \frac{-2}{\sigma^2} \int_0^1 \langle |D\phi_{t,1}^v| (I_0 \circ \phi_{t,0}^v - I_1 \circ \phi_{t,1}^v), D(I_0 \circ \phi_{1,0}^v) h_t \rangle_{L^2} dt \tag{3.27}$$

$$= \frac{-2}{\sigma^2} \int_0^1 \langle |D\phi_{t,1}^v| (J_t^0 - J_t^1), D(I_0 \circ \phi_{t,0}^v) h_t \rangle_{L^2} dt \tag{3.28}$$

$$= \frac{-2}{\sigma^2} \int_0^1 \langle |D\phi_{t,1}^v| (J_t^0 - J_t^1) \cdot \nabla J_t^0, h_t \rangle_{L^2} dt. \tag{3.29}$$

With the discussion about the dot product in vector field in Section 3.1.1, the variation of E_2 can be transposed to the vector field, which is consistent with the variation of E_1 in

Equation 3.22

$$(\partial_h E_2(v))_V = -\int_0^1 \langle K(\frac{2}{\sigma^2}|D\phi^v_{t,1}|(J^0_t - J^1_t)\cdot\nabla J^0_t), h_t\rangle_V dt. \tag{3.30}$$

Collect two terms 3.22 and 3.30, the gradient of the energy function in the vector space V at time t is given by

$$(\nabla_{v_t} E)_V = 2v_t - K\left(\frac{2}{\sigma^2}|D\phi^v_{t,1}|(J^0_t - J^1_t)\right), \tag{3.31}$$

where the subscript V indicates the gradient lives in the smooth vector space V. Set this energy gradient to 0, the Euler-Lagrange equation in Theorem 2 can be derived. □

With Theorem 2, the optimal velocity field for 3.9 is generated by standard gradient descent algorithm by using the gradient term shown in 3.31. In the next section, we will discuss the numerical implementation of this algorithm.

3.1.3 Numerical Implementation of the LDDMM Algorithm

The ODE of $\phi^v_{s,t}$ illustrates the continuous flow of transformation or diffeomorphism, we need to discretize the flow for implementation in our algorithm. The image domain is denoted as Ω Let the diffeomorphism generated by discrete flow be indexed by $t_j \in [0,1],\ j \in [0,N]$, then the size of a timestep $\delta t = \frac{1}{N}$, that is we always scale the time interval to be $[0,1]$ and assume piece-wise constant velocities in the discrete time intervals. Let our moving image $I_0 = J^0_0$ be the image at the starting point of the flow being mapped to the fixed image $I_1 = J^1_1$ at the end time point. We use the notation $v^k_{t_j}$ and $\phi^k_{t_j,0}$ for the velocity field and the transformation for the kth iteration in the gradient descent algorithm at time index t_j.

The variational optimization of the energy function (3.9) is performed in a standard gradient descent scheme

$$v^{k+1}_{t_j} = v^k_{t_j} - \epsilon\cdot\nabla_{v^k_{t_j}} E, \tag{3.32}$$

where the gradient term $\nabla_{v_{t_j}^k} E$ is acquired by discretizing the Equation 3.31 for each voxel $y \in \Omega$ as

$$\nabla_{v_{t_j}^k} E_{t_j}^k(y) = 2v_{t_j}^k(y) - K\left(\frac{2}{\sigma^2}|D\phi_{t_j,1}^v(y)|(J_{t_j}^0(y) - J_{t_j}^1(y))\right). \tag{3.33}$$

The energy for each time index t_j at kth iteration is

$$E_{t_j}^k = \delta t \cdot \sum_{j=0}^{N-1} ||v_{t_j}^k||_V^2 + \frac{1}{|\Omega|}\sum_{y\in\Omega} |J_1^0(y) - J_1^1(y)|^2 \tag{3.34}$$

$$= \delta t \cdot \sum_{j=0}^{N-1} ||Lv_{t_j}^k||_{L^2}^2 + \frac{1}{|\Omega|}\sum_{y\in\Omega} |J_1^0(y) - J_1^1(y)|^2, \tag{3.35}$$

where $|\Omega|$ is the number of pixels for the fixed or moving image, and L is the operator we have discussed in Section 3.1.1 that transforms the vector norm to standard L^2 norm.

The first term in Equation 3.35 is also regarded as the length of the path from the identity transformation Id to the matching diffeomorphism $\phi_{1,0}$ at kth iteration.

Another implementation issue needs to be resolved is the choice of operator L and K. Let discrete functions $f(x)$ and $g(x)$ be defined $f, g : \Omega \subset \mathbb{Z}^n \to \mathbb{R}^n$ on the discrete image grid $\Omega \subset \mathbb{Z}^n$ and satisfy the equation

$$L^\dagger L f = g. \tag{3.36}$$

The operator L is chosen to be of the Cauchy-Navier type, which is given by

$$L \doteq -\alpha\nabla^2 + \gamma I, \tag{3.37}$$

where I is the identity operator and $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the Laplacian operator. The parameter α enforces smoothness, higher values ensure solutions of high regularity, and the coefficient γ is chosen to be positive so that the operator is non-singular [12]. For a given discrete function $f(x, y)$, to calculate the function $g(x, y)$ via Equation 3.36

$$g(x,y) = (-\alpha\nabla^2 + \gamma I)f(x,y) \tag{3.38}$$

$$= -\alpha\left[\frac{\partial^2}{\partial x^2}f(x,y) + \frac{\partial^2}{\partial y^2}f(x,y)\right] + \gamma f(x,y). \tag{3.39}$$

Write right side term in Equation 3.39 in the discrete version using finite differences on a periodic domain.

$$-\alpha\left[\frac{f(x+\triangle x,y)+f(x-\triangle x,y)-2f(x,y)}{\triangle x^2}+\frac{f(x,y+\triangle y)+f(x,y-\triangle y)-2f(x,y)}{\triangle y^2}\right]+\gamma f(x,y) \tag{3.40}$$

With applying Fourier transformation for both sides of Equation 3.39, the equation becomes

$$G(k_1,k_2)=A(k_1,k_2)F(k_1,k_2), \tag{3.41}$$

where $A(k_1,k_2)=2\alpha\left[\frac{1-\cos(2\pi k_1\triangle x)}{\triangle x^2}+\frac{1-\cos(2\pi k_2\triangle y)}{\triangle y^2}\right]+\gamma$, $\triangle x=\frac{1}{N_1}$, $\triangle y=\frac{1}{N_2}$, N_1,N_2 is the dimension of data on x and y-axis. Here we present the proof.

Proof. Let us first apply Fourier Transform on the first term in Equation 3.39

$$\mathcal{F}\left(\frac{f(x+\triangle x,y)+f(x-\triangle x,y)-2f(x,y)}{\triangle x^2}\right)=\frac{1}{\triangle x^2}\left(e^{j\triangle x2\pi k_1}+e^{-j\triangle x2\pi k_1}-2\right)F(k_1,k_2) \tag{3.42}$$

$$=\frac{2\cos(\triangle x2\pi k_1)-2}{\triangle x^2}F(k_1,k_2). \tag{3.43}$$

Similarly, the second term after Fourier Transform is $\frac{2\cos(\triangle y2\pi k_2)-2}{\triangle y^2}F(k_1,k_2)$. Collect all the terms together, we will get Equation 3.41. □

For Equation 3.36, we assume periodic boundary conditions, in which case L is self-adjoint $L=L^\dagger$. Then we can evaluate operator $L^\dagger L$ as L^2, that yields the Fourier Transform of Equation 3.36:

$$G(k_1,k_2)=A(k_1,k_2)^2F(k_1,k_2). \tag{3.44}$$

In this way, the effect of inverse operator $K = (L^\dagger L)^{-1}$ can be expressed properly in the Fourier domain

$$F(k_1, k_2) = \frac{G(k_1, k_2)}{A(k_1, k_2)^2}, \tag{3.45}$$

$$f(x, y) = \mathcal{F}^{-1}\left(\frac{G(k_1, k_2)}{A(k_1, k_2)^2}\right) \tag{3.46}$$

$$= Kg(x, y). \tag{3.47}$$

3.1.4 LDDMM Algorithm and Experimental Results

The registration algorithm is described in following pseudo code shown in Algorithm 2. The output diffeomorphism is $\phi_{t_j,0}^{K-1}$ for $\forall j \in [0, N-1]$.

Algorithm 2 LDDMM Algorithm

1: $v_{t_j}^0 = 0$ for $\forall j \in [0, N-1]$
2: $\nabla_{v_{t_j}^0} E = 0$ for $\forall j \in [0, N-1]$
3: $\phi_{t_j,1}^k = Id$, $\phi_{t_j,k}^0 = Id$ for $\forall j \in [0, N-1]$, $k \in [0, K-1]$
4: **for** $k = 0$ to $K-1$ **do**
5: Calculate $\phi_{t_j,1}^k$ and $|D\phi_{t_j,1}^k|$ for each t_j
6: Calculate $\phi_{t_j,0}^k$ for each t_j
7: Calculate $J_{t_j}^0 = I_0 \circ \phi_{t_j,0}^k$ and $DJ_{t_j}^0$ for each t_j
8: Calculate $J_{t_j}^1 = I_1 \circ \phi_{t_j,1}^k$ for each t_j
9: Calculate the energy gradient $\nabla_{v_{t_j}^k} E$ for each t_j
10: Update new velocity field $v_{t_j}^{k+1} = v_{t_j}^k - \epsilon \cdot \nabla_{v_{t_j}^k} E$ for each t_j
11: **end for**

For each iteration, we need to integral $v_{t_j}^k$ for $\forall j \in [0, N-1]$ to generate the diffeomorphism $\phi_{t_j}^k$, this function is shown in following Algorithm 3.

Algorithm 3 Calculate Diffeomorphism from Velocity Field

$\phi^k_{t_j,0} = Id$
$\phi^k_{1,t_j} = Id$
for $j = 0$ to $N-1$ **do**
  $\phi^k_{t_{j+1},0} = \frac{1}{N} \cdot \phi^k_{t_j,0} \circ v^k_{t_j} + \phi^k_{t_j,0}$
  $\phi^k_{t_{j+1},1} = \frac{1}{N} \cdot D\phi^k_{t_j,1} \cdot v^k_{t_j} + \phi^k_{t_j,1}$
end for

Parameters should be carefully selected, the following Table 3.1 shows the effect of changing some parameters and values we used in our implementation. The parameters set up may not be the optimal but is the best among our experiments.

Table 3.1: Parameters Set Up for LDDMM

Parameter Name	**Description**	**Value**
α in L, K operator	Enforce smoothness to the vector field	1.5
σ in the energy function	Weight coefficient for image metric	0.06
N	Number of discrete time intervals between $[0, 1]$	50
K	Maximum number of interactions for gradient descent	50
ϵ	Fixed step size for gradient descent	0.1

We first present the results of implementing our LDDMM algorithm to META data set. For all sources from META, they are gray images obtained from the same device, which provide a better condition for registration comparing with multi-modal OMEGA data.

We randomly chose three images from the data set and use one of them to be the moving image, others will be treated as fixed images. Results are shown below.

As we can see in Figure 3.1, the moving image changes dramatically to the fixed image, with the error (we are using SSD to measure the difference between two images) starting from 7486 to 135. The processes of convergence of energy function and SSD error are shown in Figure 3.2.

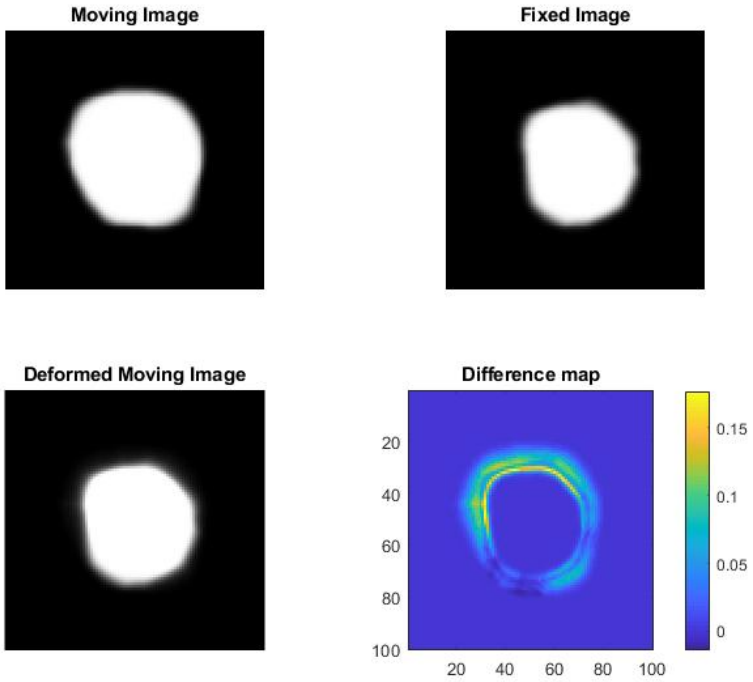

Figure 3.1: Registration Result 1 for META data

The flow of diffeomorphisms is shown in Figure 3.3, with the metric that enforces smoothness constraint on the vector field, the LDDMM algorithm generates a differentiable and invertible transformation.

We could see the property of a differentiable transformation from the above figure, the moving image I_0 starts at time $t = 0$ and gradually deforms over discrete time $(t_j,\ j = 13, 25, 37)$ till the endpoint $t = 1$. This sequence of images shows the path of how diffeomorphisms moving on a smooth and differentiable manifold. Actually, Figure 3.3 shows image I_0 composed with the flow of diffeomorphisms, but we can use this to visualize the process happens in the manifold of diffeomorphisms. The moving image at time $t = 1 :\ I_0 \circ \phi_{1,0}$ is the final registration result to match the fixed image I_1. To justify our algorithm, we will present another registration result below.

Figure 3.4 and 3.5 give registration results for another pair of images. Similar to the discussion above, the LDDMM algorithm has a robust performance for the singular source image.

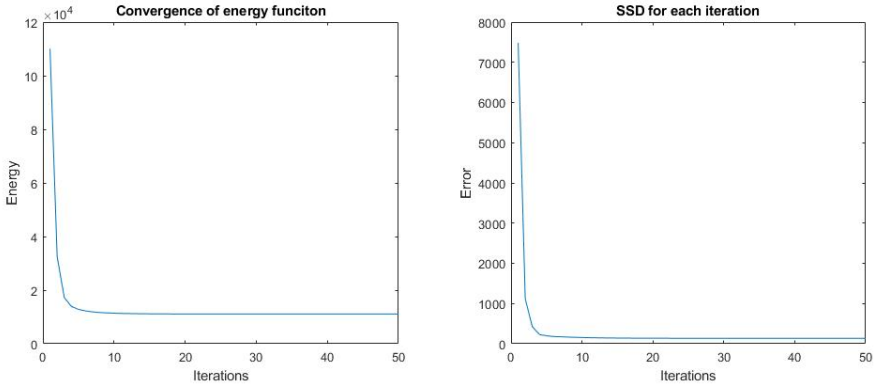

Figure 3.2: Value of Energy Function and SSD Error for each iteration

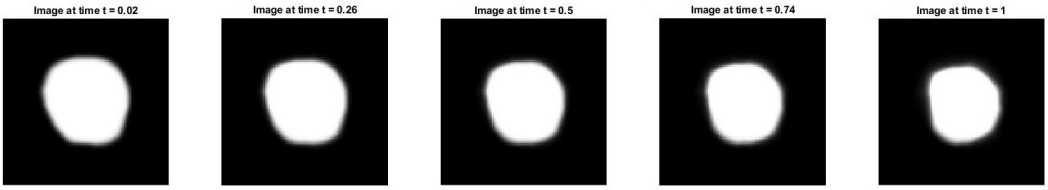

Figure 3.3: Flow of diffeomorphisms

However, when applying the same method to hyperspectral imagery, the same algorithm becomes unstable and even fail for most cases. Here we present the result of implementation with OMEGA in Figure 3.6.

The energy function in the left part of Figure 3.6 has an unexpected peak for the first and second iteration. Even though it starts to show the trend of convergence after the fifth iteration, the value of energy fluctuates a lot and stops at a level with a higher value than the original one. Because of the multi-modal issue, SSD error no longer makes sense for comparing two hyperspectral images. Thus, we borrow the idea of mutual information in Section 2.1.2 to quantify the distance between two images from OMEGA data and the function of mutual information subjects to iterations is shown in the right part of Figure 3.6. The mutual information function with iterations has the same shape with the energy function, still does not converge. To show how applying transformation affects the mutual information, we present some experiments of slightly moving image grid for two directions and compute the mutual information between them in Appendix A.

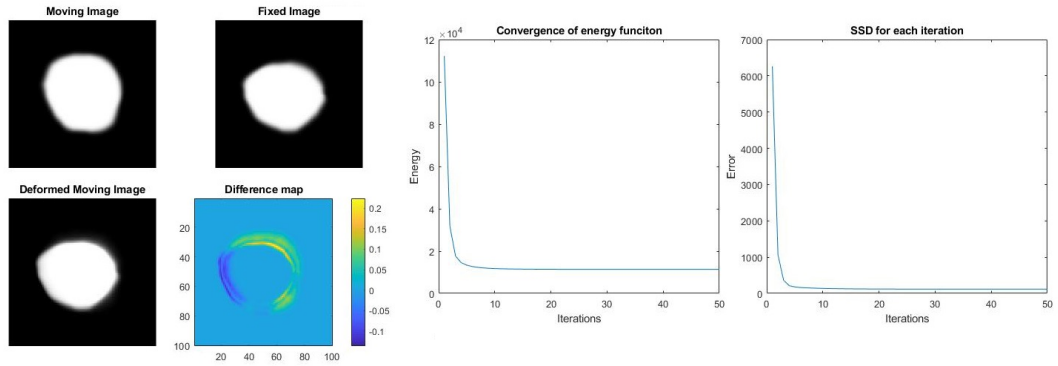

Figure 3.4: Registration Result 2 for META data

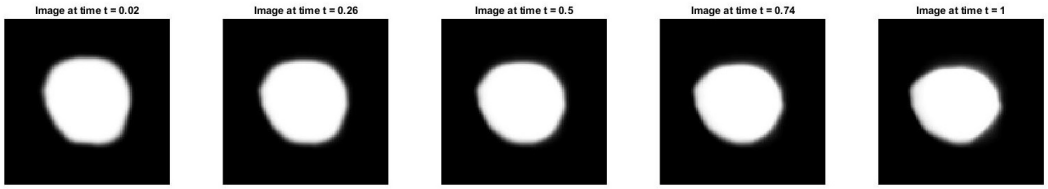

Figure 3.5: Flow of diffeomorphisms

Because images in the OMEGA data set are too large to display on paper, we only extract some critical parts with abundant feature information, such as meteor craters or boundaries of mountain chains to show the result. Figure 3.7 shows three critical areas (Feature 1 shows the area from row 1025 to 1065, Feature 2 is 2030 to 2065 and Feature 3 is 702 to 742) for the registration result. The average error is 2 pixels on y-axis and the mutual information after registration is only around 1.7, which is not a satisfactory result.

In the following section, a modified LDDMM algorithm will be presented to improve the registration performance for hyperspectral images. The core idea of our method is to augment the current energy function with another matching term that includes boundary information.

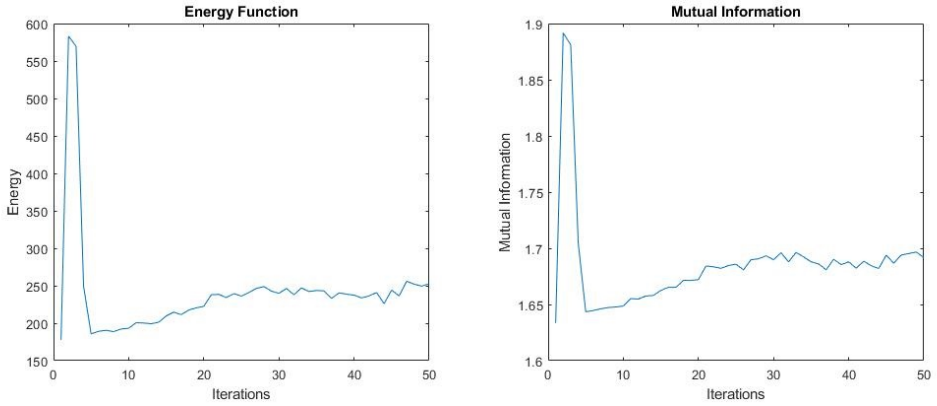

Figure 3.6: Registration Result for applying LDDMM to OMEGA data

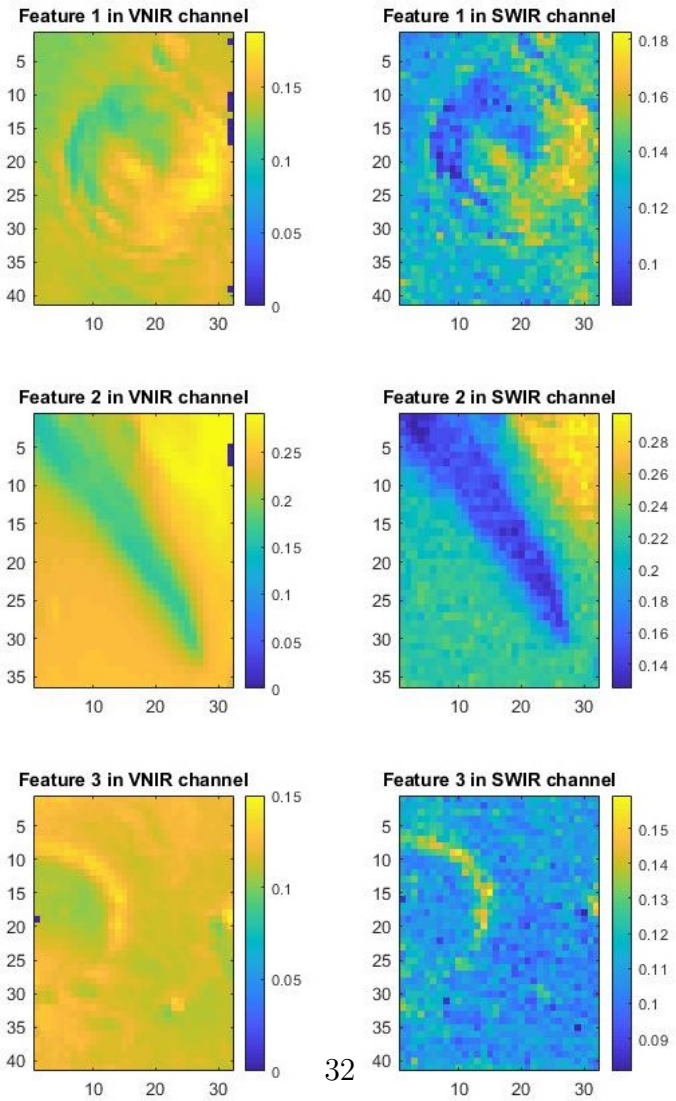

Figure 3.7: Registration Result for Critical Areas

3.2 Modified LDDMM with Edge Information

In the previous Section 3.1, we have thoroughly discussed the LDDMM algorithm and tried to implement it to our hyperspectral image data set OMEGA. However, the results are not satisfactory. The reason for the failure of the LDDMM algorithm is because of the SSD distance we have used in the energy function. For multi-modal imagery, even two images are perfectly aligned, the sum of squared distance is still likely to be large. Therefore, SSD cannot be the only term to decide the "goodness" of matching result. In our modified LDDMM, we are adding an edge matching term in the energy function that compares the distance of edge maps for moving and fixed images. We call this new energy function as augmented energy.

3.2.1 Edge Detection

Edge detection includes a variety of mathematical methods that aim at identifying points in an image at which the image brightness changes sharply or, more formally, has discontinuities. The points at which image intensity changes dramatically are typically organized into a set of curved line segments termed *edges* [19].

Before we start discussing about the edge detection method, we want to first emphasize why edge map will help register hyperspectral images. OMEGA is using bi-dimensional charge coupled device (CCD) to measure the intensity of input light for different spectral. The simple mathematical model of sensor is given by

$$I^0[n_x, n_y, \lambda] = \int_{t=0}^{T} \left[\int E(\lambda, x, y, t) p(x - x^-_{n_x}, y - y^-_{n_y}) q(\lambda) d\lambda dx dy \right] dt, \tag{3.48}$$

$$I[n_x, n_y, \lambda] \doteq I^0[n_x, n_y, \lambda] + \epsilon(\lambda). \tag{3.49}$$

where $E(\lambda, x, y, t)$ is the light energy per unit area per unit time per spectral, arriving at time point (x, y) at time t. $I^0[n_x, n_y]$ is the intensity measured by the sensor at grid location $[n_x, n_y]$, $n_x, n_y \in \mathbb{N}$. $p(x - x^-_{n_x}, y - y^-_{n_y})$ is the sensitivity function, for the ideal case is 1 inside pixel and 0 outside. $q(\lambda)$ is the quantum efficiency of the sensor: Ratio of light energy

to charge/voltage, is a function to light wave length λ. ϵ is the signal-independent addition noise, most time can be treated as Gaussian noise.

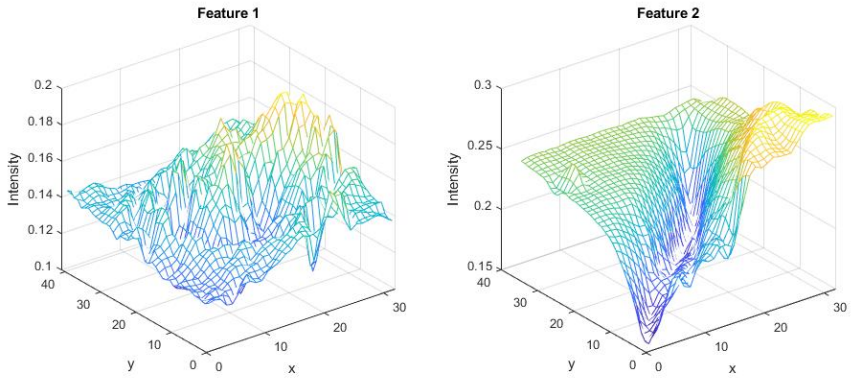

Figure 3.8: 3D Visualization of Intensity Value in SWIR channel

Figure 3.8 shows two areas within the same band in SWIR channel, there are some parts with discontinuity, which may be treated as edges. We need to figure out if all those edges represent the geometry features of the Mars surface that remain the same when we acquired data. If the answer is yes, then using edge maps in registration will provide some useful information for matching two hyperspectral images.

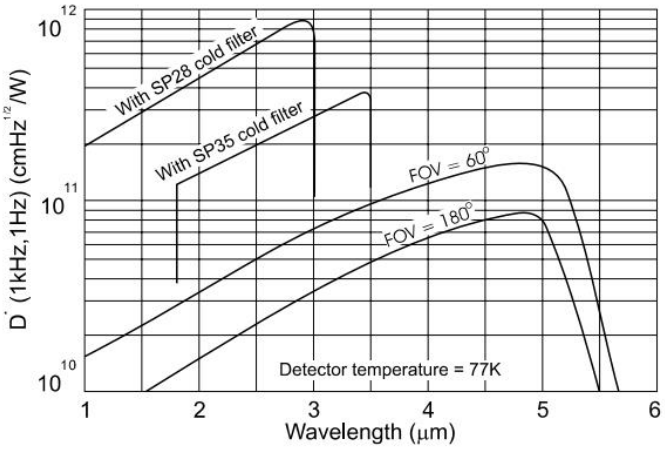

Figure 3.9: Quantum Efficiency for Detector in SWIR Channel

Figure 3.9 [15] shows the quantum efficiency for the detector (InSb linear array) in the SWIR channel. It is obvious to find that the efficiency (reflectance) function is continuous among the measuring spectral range from $0.93\mu m - 5.2\mu m$. Therefore, the discontinuity of intensity value from Figure 3.8 must come from the light energy function $E(\lambda, x, y, t)$ with the assumption that the sensitivity function p is ideal.

When discussing the light energy function, we need to introduce a light reflection model. Mars surface will absorb and reflect portions of the incident light coming from Sun, instruments on OMEGA then measure the radiance of reflected light from a certain viewpoint on the orbit. A general view of this process is shown in Figure 3.10.

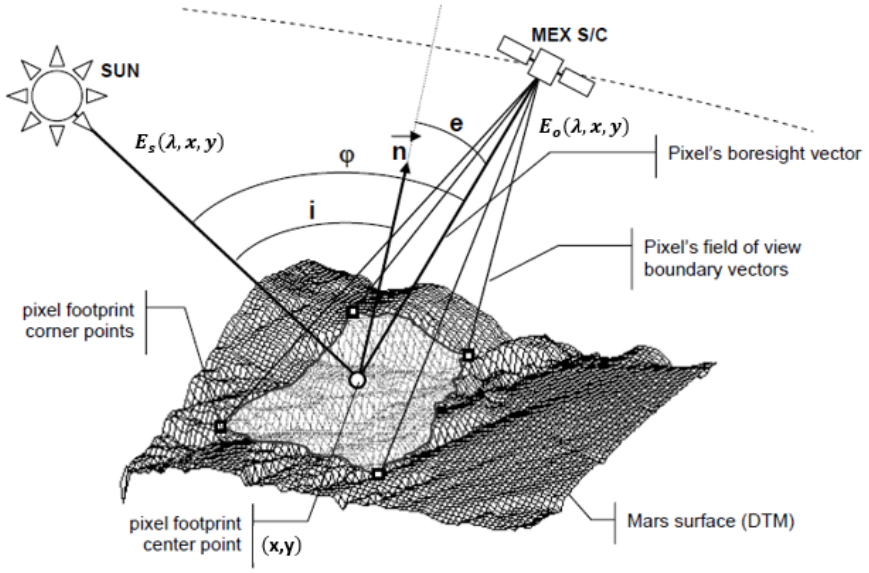

Figure 3.10: Overview of Pixel Observation Geometry

The relationship between input radiation $E_s(\lambda, x, y)$ and output radiation $E_o(\lambda, x, y)$ at location (x, y) is given by

$$E_o(\lambda, x, y) = q(\lambda, x, y, i, e, \vec{n}) E_s(\lambda, x, y). \tag{3.50}$$

All incident and reflection angles i, e are recorded. Given that, the function $q(\lambda, x, y, i, e, \vec{n})$ only depends on surface geometry and mineral composition, which involves discontinuity when the surface has some dramatic shape or mineral changes. Under this assumption, edge

maps that extract discontinuous parts in hyperspectral images will provide strong information about the underlying geometry of surface that can be regarded as a constant. Additional noise will also generate discontinuity, but with proper denoising method in edge detection algorithm, we could significantly reduce the effect of noise.

Here we present some edge maps to show that they share the same landmark features even though the sources are coming from different instruments in Figure 3.11.

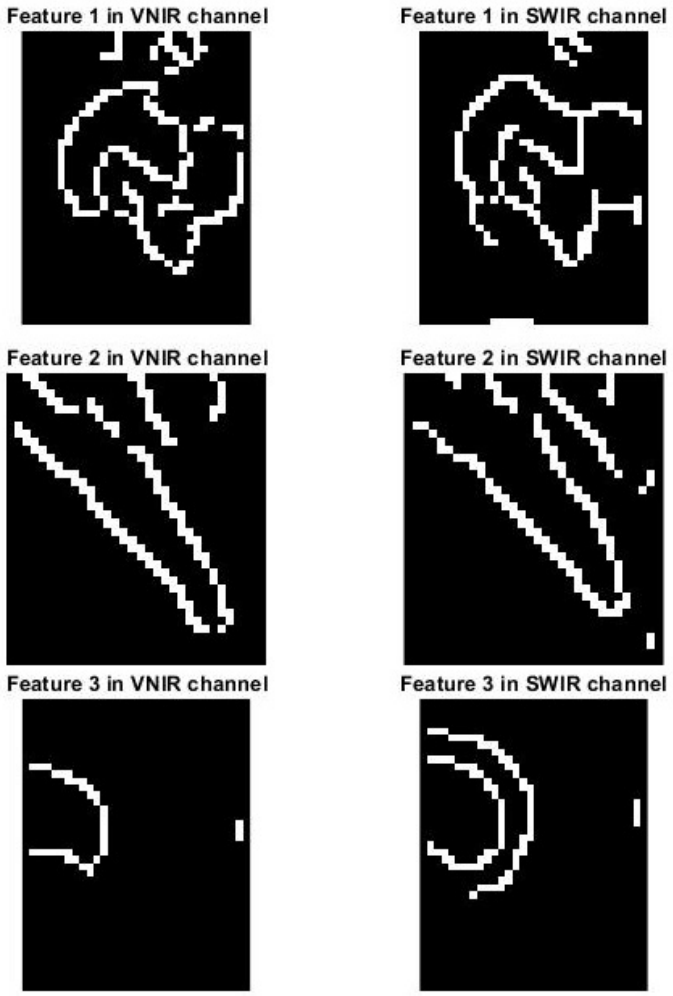

Figure 3.11: Edge Maps for SWIR and VNIR Channel

Edge maps in Figure 3.11 are obtained through the Canny edge detector, which is a mature method to generate a clear picture of edges. We will only summary the process of Canny

edge detection briefly in this paper, but details can be found at [7]. The process of the Canny edge detection algorithm can be broken down to 5 different steps:

1. Apply Gaussian filter to smooth the image in order to remove some noise.
2. Find the intensity gradients of the image.
3. Apply non-maximum suppression to get rid of spurious response to edge detection.
4. Apply double threshold to determine potential edges
5. Track edge by hysteresis: Finalize the detection of edges by suppressing all the other edges that are weak and not connected to strong edges.

3.2.2 Variational Minimization of the Augmented Energy Function

With the discussion about how edge maps contain some useful geometric information in Section 3.2.1, we want to further extend this idea to registration algorithm, specifically for LDDMM. The original LDDMM only use the SSD to evaluate the difference between two images, which is not a good criterion for hyperspectral imagery. In this section, we will discuss our modified LDDMM algorithm with additional edge difference penalty.

The modified LDDMM algorithm derives from the modification of the original energy function. We are adding an edge matching term as a penalty function. The augmented energy function is given as

$$E_{\text{aug}}(v) = \underbrace{\int_0^1 ||v_t||_V^2 dt}_{E_1(v)} + \underbrace{\frac{1}{\sigma_1^2}||I_0 \circ \phi_{1,0}^v - I_1||_{L^2}^2}_{E_2(v)} + \underbrace{\frac{1}{\sigma_2^2}||E[I_0] \circ \phi_{1,0}^v - E[I_1]||_{L^2}^2}_{E_3(v)}, \tag{3.51}$$

where $E[\cdot]$ is the Canny edge operator.

Before calculating the gradient of augmented energy function subject to velocity field, we first give an intuitive gradient constancy assumption.

Theorem 3. *The gradient of the image intensity value can be assumed not to vary due to the displacement u [20]. This gives*

$$\nabla I(x, y, t) = \nabla I(x + u_x, y + u_y, t + 1). \tag{3.52}$$

In our case, with the notation in 3.1.1, Equation 3.52 can be rewritten as

$$\nabla I(x, y) = \nabla (I \circ \phi^v_{1,0})(x, y). \tag{3.53}$$

With this assumption, we can derive an approximation of how transformation/diffeomorphism affects edge maps. This is given by

$$E[I_0] \circ \phi^v_{1,0} \approx E[I_0 \circ \phi^v_{1,0}]. \tag{3.54}$$

This approximation makes sense based on the methodology of the Canny edge detection algorithm. Because it mainly uses the intensity gradient of the image to find the center of edges as we discussed in Section 3.2.1.

The reason we are using this approximation of the deformed edge map is based on the interpolation method in our algorithm. For those transformations that generate the displacement field with non-integer values, the deformed image grid will have some non-integer indexes. For these cases, our algorithm applies a bilinear interpolation method to estimate values for these non-integer index pixel.

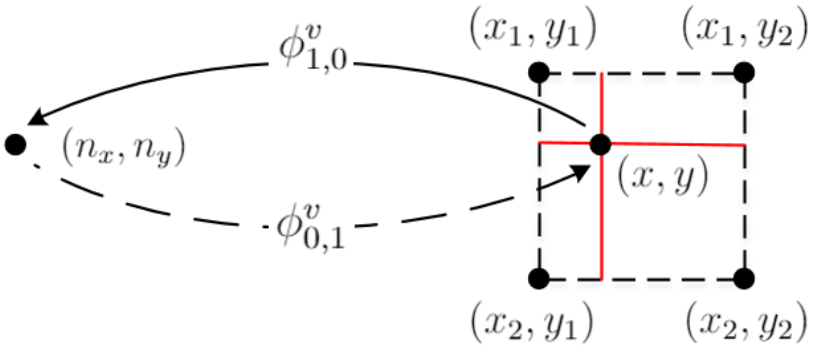

Figure 3.12: Bilinear Interpolation

In Figure 3.12, (n_x, n_y) is an index in the deformed image with corresponding coordinates (x, y) in the original image. The value $g(n_x, n_y)$ is given by

$$g(n_x, n_y) = f(x, y) \tag{3.55}$$

$$= \frac{1}{(x_2 - x_1)(y_2 - y_1)} \begin{bmatrix} x_2 - x & x - x_1 \end{bmatrix} \begin{bmatrix} f(x_1, y_1) & f(x_1, y_2) \\ f(x_2, y_1) & f(x_2, y_2) \end{bmatrix} \begin{bmatrix} y_2 - y \\ y - y_1 \end{bmatrix}, \tag{3.56}$$

where $x_1 = \lfloor x \rfloor, x_2 = \lceil x \rceil$ and $y_1 = \lfloor y \rfloor, y_2 = \lceil y \rceil$.

Here we present some experimental results in Figure 3.13 to show this approximation. The left column in Figure 3.13 is the edge map for $E[I_0] \circ \phi^v_{1,0}$ and the right column is generated by $E[I_0 \circ \phi^v_{1,0}]$. It is obvious that the left one has lots of blur at the boundary which violates the general criteria for edge detection: the edge point detected from the operator should accurately localize on the center of the edge. However, edge maps from our approximation give a clear image and will also show those pixels that are moved out of the image domain. This will give better robustness for our registration algorithm.

Theorem 4. *The variation of $E_3(v)$ is given by*

$$\partial_h E_3(v) = -\frac{2}{\sigma_2^2} \int_0^1 \langle |D\phi^v_{t,1}| (EJ^0_t - EJ^1_t) \cdot \nabla EJ^0_t, h_t \rangle_{L^2} dt, \tag{3.57}$$

where $EJ^0_t \doteq E[I_0 \circ \phi_{t,0}]$, $EJ^1_t \doteq E[I_1 \circ \phi_{t,1}]$.

Proof. The Let the velocity $v \in L^2([0,1], V)$ be perturbed by an ϵ amount direction $h \in L^2([0,1], V)$. The variation $\partial_h E(v)$ of the energy function is defined by

$$\partial_h E_3(v) = \lim_{\epsilon \to 0} \frac{\partial E_3(v + \epsilon h)}{\partial h} \tag{3.58}$$

Similar with Equation 3.24, the variation of $E_3(v)$ is given by

$$\partial_h E_3(v) = \frac{2}{\sigma_2^2} \langle E[I_0] \circ \phi^v_{1,0} - E[I_1], DE[I_0] \circ \phi^v_{1,0} \partial_h \phi^v_{1,0} \rangle_{L^2}. \tag{3.59}$$

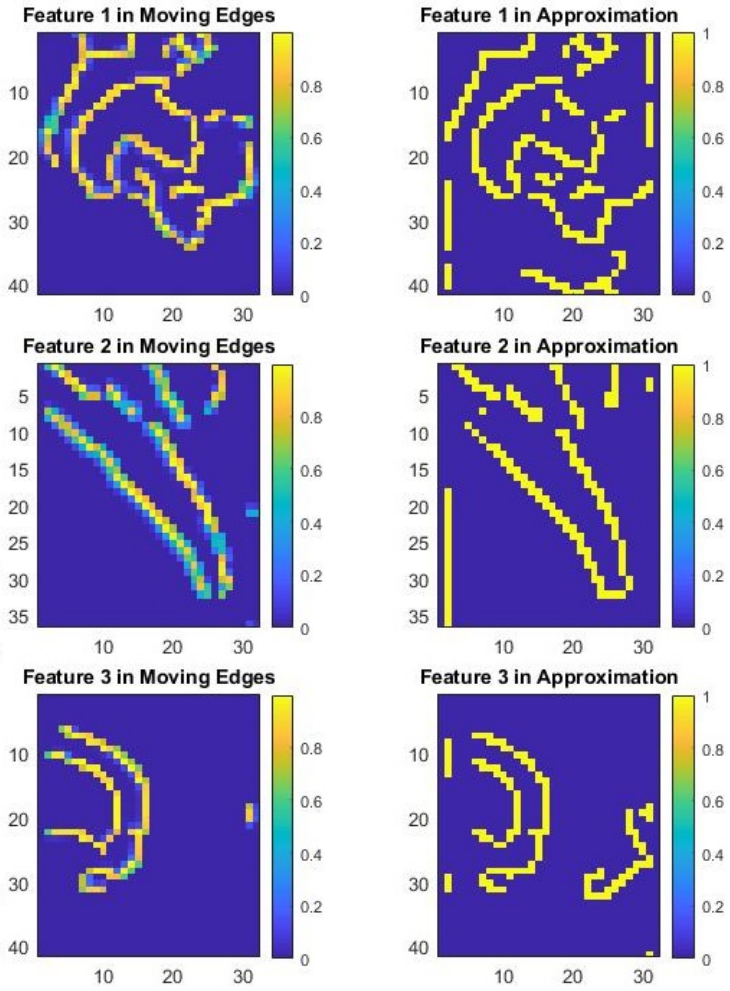

Figure 3.13: Edge Maps for Different Deformation Methods

Following the same procedure in Theorem 2, the variation becomes

$$\partial_h E_3(v) = \frac{-2}{\sigma_2^2} \int_0^1 \langle |D\phi_{t,1}^v| (E[I_0] \circ \phi_{t,0}^v - E[I_1] \circ \phi_{t,1}^v), D(E[I_0] \circ \phi_{1,0}^v) h_t \rangle_{L^2} dt \tag{3.60}$$

$$= \frac{-2}{\sigma^2} \int_0^1 \langle |D\phi_{t,1}^v| (E[I_0] \circ \phi_{t,0}^v - E[I_1] \circ \phi_{t,1}^v)) \cdot \nabla (E[I_0] \circ \phi_{t,0}^v), h_t \rangle_{L^2} dt. \tag{3.61}$$

Then substitute Equation 3.54 to Equation 3.61, we will get Equation 3.57. □

By collecting all variations for $E_1(v)$, $E_2(v)$ and $E_3(v)$, the gradient of augmented energy function is given by

$$(\nabla_{v_t} E_{\text{aug}})_V = 2v_t - K\left(\frac{2}{\sigma_1^2}|D\phi_{t,1}^v|(J_t^0 - J_t^1)\right) - K\left(\frac{2}{\sigma_2^2}|D\phi_{t,1}^v|(EJ_t^0 - EJ_t^1)\cdot\nabla EJ_t^0\right). \quad (3.62)$$

The optimal vector field $\hat{v}_t$ is obtained by using the gradient in Equation 3.62 in a standard gradient descent algorithm. Follow the same structure of algorithm shown in Section 3.1.4, our modified large deformation diffeomorphic metric mapping (LDDMM) is finished. We will present our results in the next section with the comparison with other methods.

3.2.3 Experiment Results

Setting up suitable coefficients in our modified LDDMM algorithm is important, here we present values we used to generate the results shown in this section.

Table 3.2: Parameters Set Up for Modified LDDMM

Parameter Name	**Description**	**Value**
α in L, K operator	Enforce smoothness to the vector field	2
σ_1 in energy function	Weight coefficient for image metric	0.08
σ_2 in energy function	Weight coefficient for edge map	0.2
N	Number of discrete time intervals between $[0, 1]$	80
K	Maximum number of interactions for gradient descent	60
ϵ	Fixed step size for gradient descent	0.1
λ in edge detector	Standard deviation of the Gaussian filter	0.2

With the parameters shown in Table 3.2, the function of energy with iterations are shown in the left part of Figure 3.14 and how mutual information increases with iterations is shown in the right part. Comparing with Figure 3.6, it is obvious that our energy function and mutual information converge nicely to a better level. The values for modified LDDMM energy function are not compatible with the previous one, for we are using different weight coefficients and adding another penalty term. However, the shape of the energy function still illustrates our augmented energy function has less relative minimums that make the gradient descent algorithm more robust than original LDDMM. The plot for mutual information in our algorithm converges to a much higher value of 1.92 compared with 1.67 in the previous method.

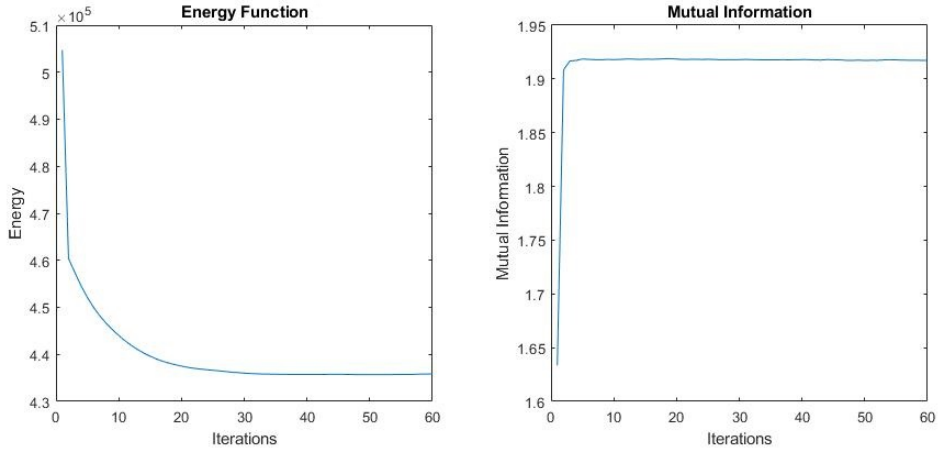

Figure 3.14: Convergence of Energy Function and MI for Modified LDDMM

The displacement fields for our modified algorithm and the original one are shown in Figure 3.15. The displacement field from our modified method gives a transformation that tends to move the pixels in the No.3 feature area in the direction $(1, -1)$ (displacement field shown in Figure 3.15 is the inverse transformation) which is consistent with the result from Appendix A (the direction is $(0, -2)$). However, the result from the original LDDMM is close to zero vector, which means there is actually no deformation for this part of the area.

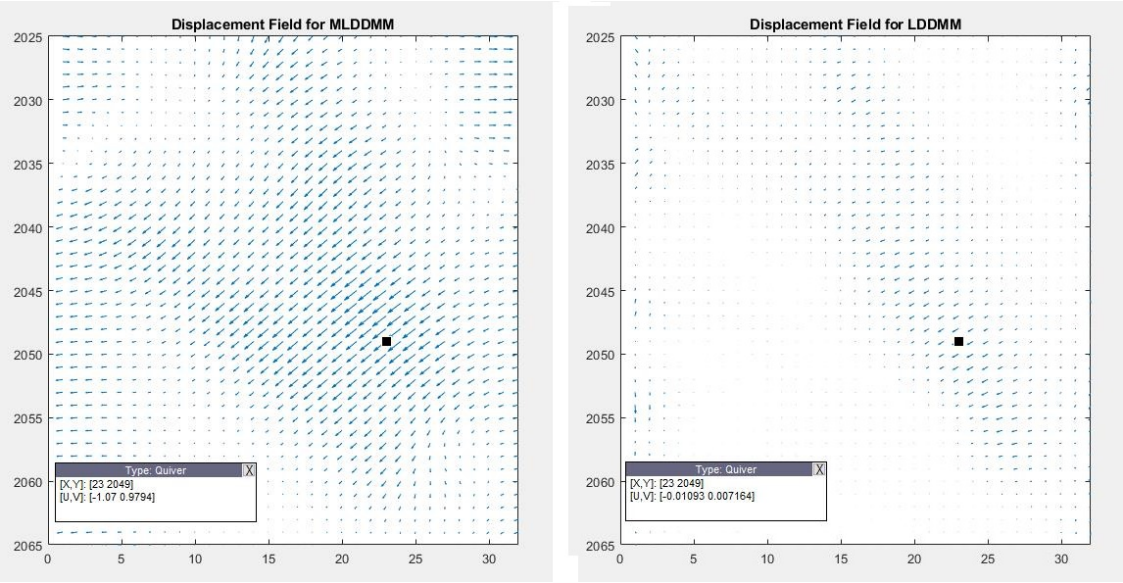

Figure 3.15: Comparison of Two Methods for Displacement Field

To test the robustness of our method, different pair of bands will be registered by our MLDDMM algorithm. Current result of pair (No.53 in VNIR channel and No.5 in SWIR channel) is shown above. Some results for other pairs are shown below.

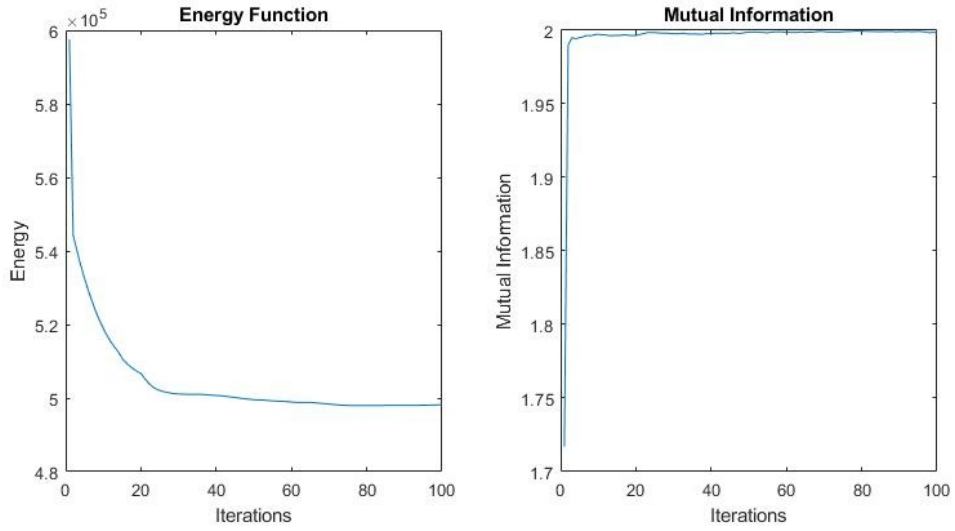

Figure 3.16: Registration Result for Band 34 and 4

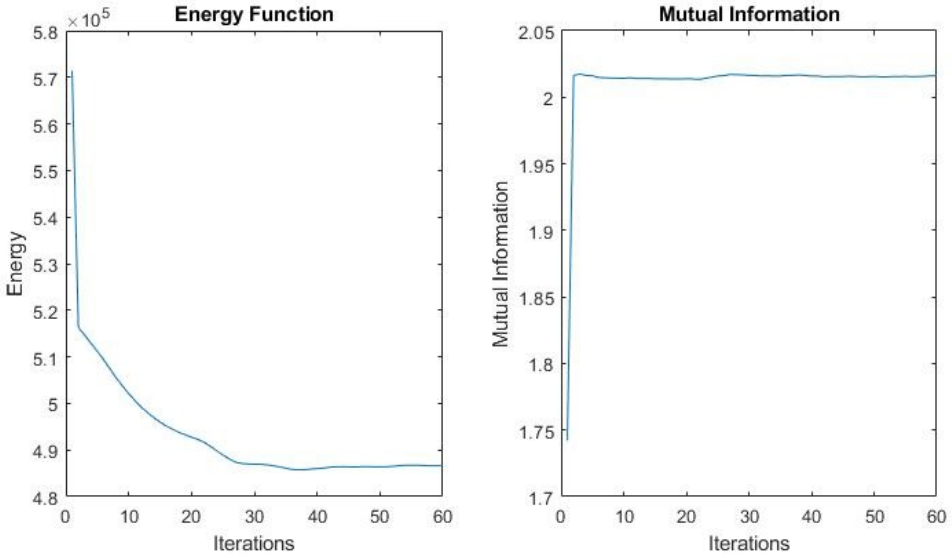

Figure 3.17: Registration Result for Band 28 and 5

The energy function and the mutual information in Figure 3.16 and 3.17 also converge to certain values but takes more iterations. We will also compare the transformation vector fields generated by each pair in Figure 3.18. The left figure is the vector field we have shown before and the other two are pair (28 and 5, 34 and 4). Slight difference occurs for each pair but they still share a common feature around the center area. Comparing those results for both convergence of the energy function and the mutual information and the displacement fields, our modified LDDMM algorithm shows the property of robustness.

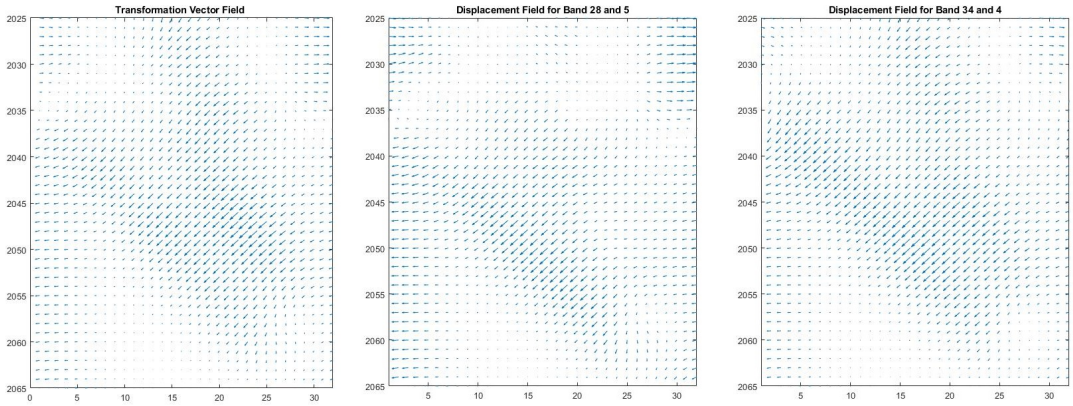

Figure 3.18: Displacement Vector Fields for Different Pair of Bands

To better visualize our result, Figure 3.19 and 3.20 shows two critical areas with easily recognized landmarks. For each area, the plots on the left top and right bottom corner are the fixed image from SWIR channel. The left bottom figure is the original moving image from the VNIR channel and our registration result is presented on the top right corner. The red reference lines are added to these figures to explain the alignment for vertical and horizontal directions.

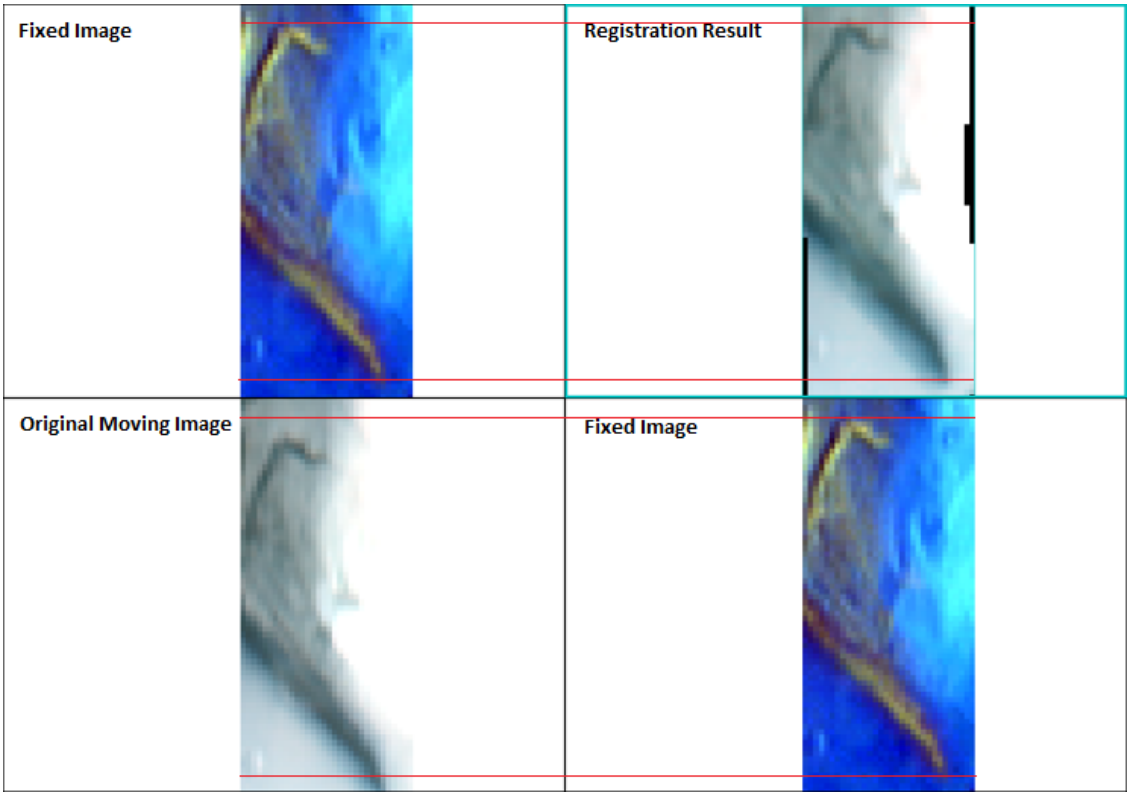

Figure 3.19: Registration Result 1 for Modified LDDMM on OMEGA

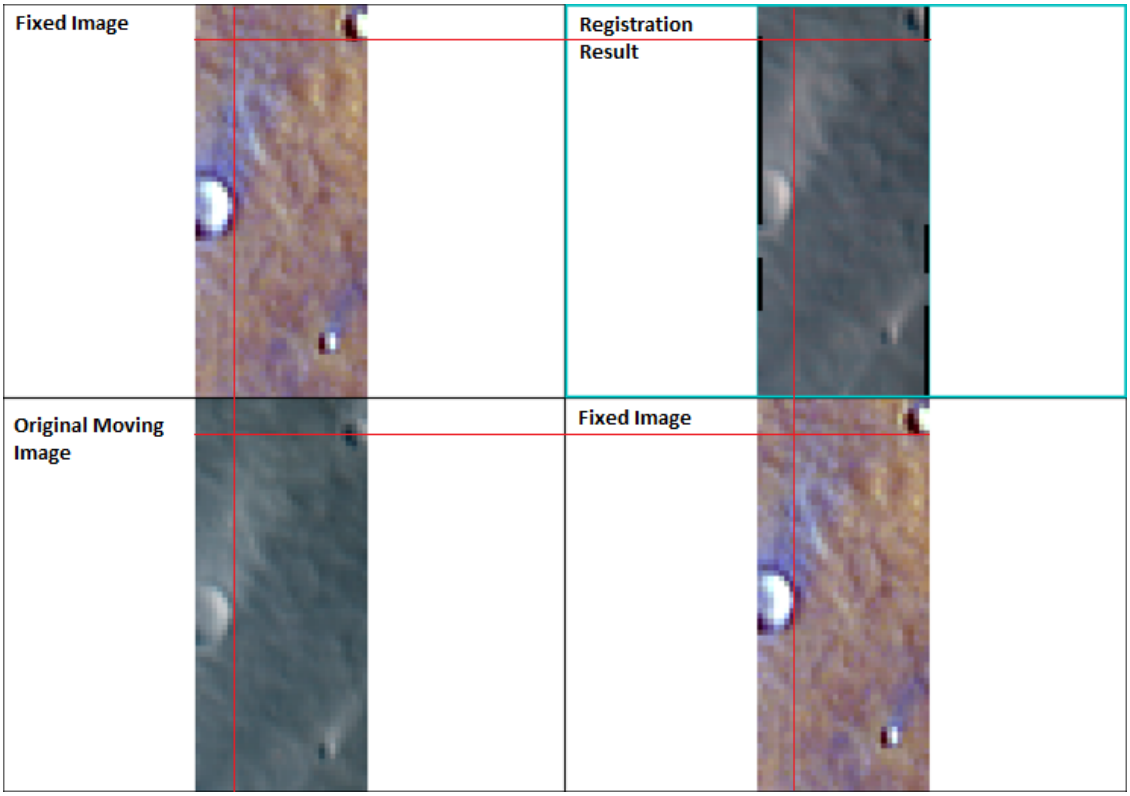

Figure 3.20: Registration Result 2 for Modified LDDMM on OMEGA

Chapter 4

Conclusion and Future Work

This thesis has attempted to develop a modification for the LDDMM algorithm in order to deal with the deterioration when applied to hyperspectral images. In Chapter 2, we describe the mutual information-based band selection algorithm to reduce the data complexity. Then in Chapter 3, this thesis first goes over the math detail for the LDDMM algorithm and then presents our modification with augmented energy function set up. By involving edge information, s better matching result comparing with the result from LDDMM is achieved.

Future work on registration may involve the following issues this thesis did not resolve. The first thing is the computational complexity. All discrete velocity fields are generated independently by the gradient of the energy function, that is when we are using 50 discrete time intervals for each iteration, the gradient needs to be computed 50 times. And calculation of the energy gradient is time-consuming for the implementation of the K operator (even in Fourier domain). However, recently, a fast approach to generate all velocity field is proposed as geodesic shooting. The idea of this method is to obtain the ODE of the flow of diffeomorphisms that the final transformation is only determined by the initial velocity field v_0. This geodesic shooting approach is discussed in Appendix B. The unsolved part for this method is to calculate the energy gradient $\nabla_{v_0} E_k$ subject to v_0 for each iteration k. Also, the choice of coefficients in our algorithm is made by experiments and will vary a lot when given different inputs.

Appendix A

Influence for Transformation on Mutual Information

In Section 3.1.4, we introduced a mutual information criterion to determine the distance between two hyperspectral images. The definition is given by

$$MI(I_0; I_1) = H(p(I_0)) + H(p(I_1)) - H\left(p(I_0), p(I_1)\right) \tag{A.1}$$

where $p(I_0)$ and $p(I_1)$ are the marginal distributions (histograms) of image I_0 and I_1 and H represents entropy. Mutual information evaluates the dependence of two distribution that merely depends on intensity values. To illustrate mutual information can still work as a criterion for image registration result, we are testing the influence of applying a transformation on mutual information for hyperspectral images.

For a given band from the OMEGA data set, we apply different levels (-15 to 15 in pixel size) of translations, and calculate the mutual information between the original image and the translated image. The result is shown below. In Figure A.1, the mutual information is monotonically decreasing as the translation becomes more severe for both directions and both images in two channels. The global maximum appears at the origin $(0, 0)$, when no translation is added. In this case, using mutual information to evaluate the difference is practicable.

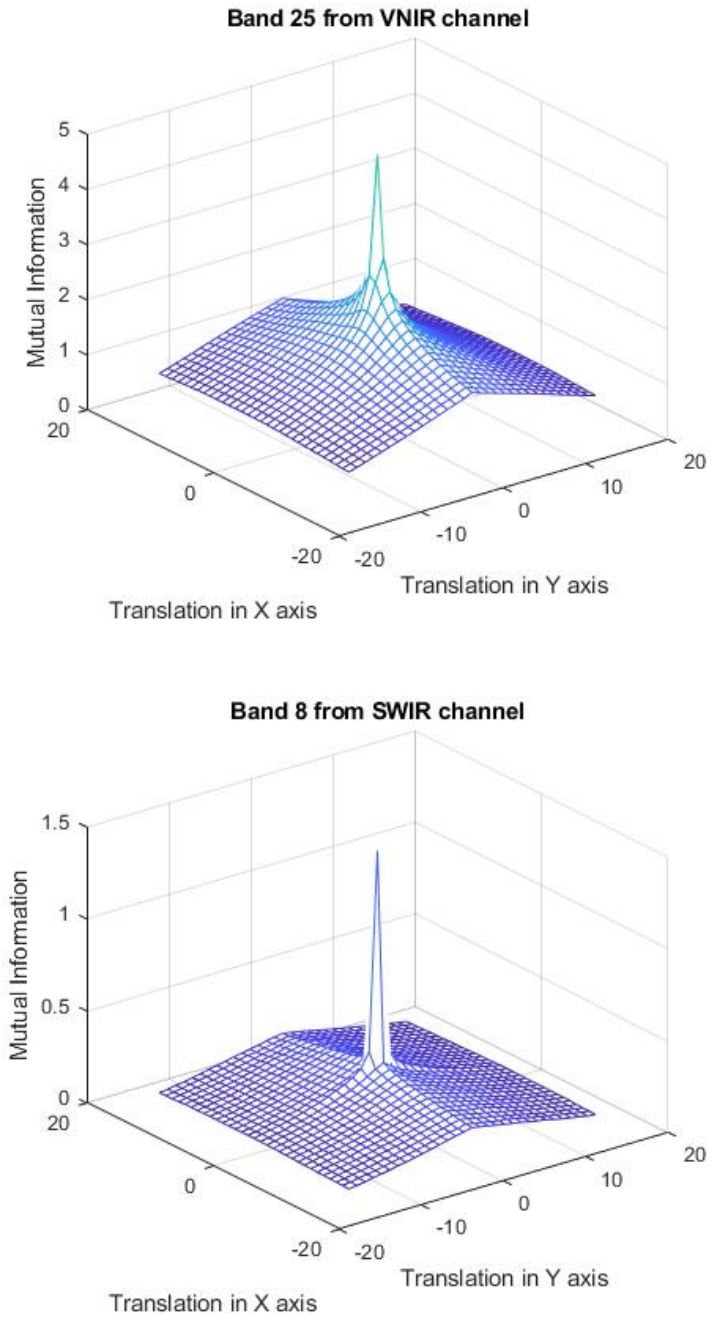

Figure A.1: Applying Different Levels of Transformation

An interesting result is shown in Figure A.2, when we apply the same procedure to the two selected images from each channel. The mutual information reaches the highest value at location $(0, -2)$, which means if we are only using translations to register these two images, applying a translation for -2 pixel size on the y-axis is the best choice. This is similar to our result in Section 3.2, however, our modified LDDMM algorithm gives back a higher value of mutual information at 1.92.

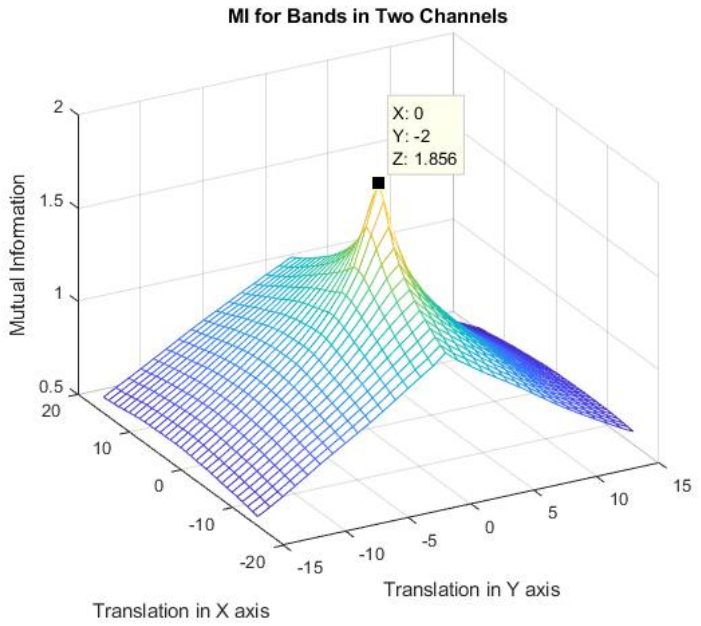

Figure A.2: Mutual Information with Different Transformations

Appendix B

Geodesic Shooting

We first need to define "straight lines" in the manifold that our diffeomorphisms live. For two given transformation ψ_0 and ψ_1, a straight line or a geodesic is the path that requires minimum energy for ψ_0 to convert to ψ_1. For our registration problem, our diffeomorphisms always start at identity transformation $\psi_0 = Id$. This energy used to evaluate the distance between Id and ψ_1 is denoted as kinetic energy, given by

$$d(Id, \psi_1)_V = \int_0^1 ||v_t||_V^2 dt \Big|_{\phi_1^v = \psi_1} \tag{B.1}$$

According to Equation 3.6, the vector norm can be converted to L_2 norm

$$d(Id, \psi_1)_{L^2} = \int_0^1 (Lv_t, v_t)^2 dt \Big|_{\phi_1^v = \psi_1} \tag{B.2}$$

Theorem 5. *The Euler equation for the kinetic energy associated to the construction of paths with minimal kinetic energy between two given points is given by:*

$$\frac{dLv}{dt} + ad_v^*(Lv) = 0 \tag{B.3}$$

Proof. Let $(t \to \phi_t)$ be extremal and $((t, \epsilon) \to \phi_{t,\epsilon})$ be a smooth deformation around $\epsilon = 0$, for convenience, denote $\phi_{t,0} = \phi_t$. Denote $\frac{\partial \phi_{t,\epsilon}}{\partial t} = v_{t,\epsilon} \circ \phi_{t,\epsilon}$, $\frac{\partial \phi_{t,\epsilon}}{\partial \epsilon} = \eta_{t,\epsilon} \circ \phi_{t,\epsilon}$, $\frac{\partial v_{t,\epsilon}}{\partial \epsilon} = h_{t,\epsilon}$, still for convenience, denoting $v_{t,0} = v_t$, $\eta_{t,0} = \eta_t$, $h_{t,0} = h_t$.

Firstly, we need to express h_t in function of other variables. Variables should satisfy following equation:

$$\frac{\partial^2 \phi}{\partial \epsilon \partial t} = \frac{\partial^2 \phi}{\partial t \partial \epsilon} \tag{B.4}$$

which yields

$$\frac{\partial \eta_t}{\partial t} \circ \phi_t + d_{\phi_t} \eta_t v_t \circ \phi_t = h_t \circ \phi_t + d_{\phi_t} v_t \eta_t \circ \phi_t \tag{B.5}$$

applying ϕ_t^{-1} on the right to both terms gives

$$h_t = \frac{\partial \eta_t}{\partial t} + d\eta_t v_t - dv_t \eta_t = \frac{\partial \eta_t}{\partial t} + [\eta_t, v_t] = \frac{\partial \eta_t}{\partial t} - ad_{v_t} \eta_t \tag{B.6}$$

The first variation of the energy is given by

$$\frac{d}{d\epsilon} \int_0^1 ||v_{t,\epsilon}||_V^2 dt \Big|_{\epsilon=0} = \frac{d}{d\epsilon} \int_0^1 \langle v_t, v_t \rangle_V^2 dt \tag{B.7}$$

$$= 2 \cdot \int_0^1 2 \cdot \langle v_t, h_t \rangle_V dt \tag{B.8}$$

$$= 4 \cdot \int_0^1 (Lv_t, h_t) dt \tag{B.9}$$

$$= 4 \int_0^1 (Lv_t, \frac{\partial \eta_t}{\partial t} - ad_{v_t} \eta_t) dt \tag{B.10}$$

$$= 4 \int_0^1 (Lv_t, \frac{\partial \eta_t}{\partial t}) dt - 4 \int_0^1 (ad_{v_t}^* Lv_t, \eta_t) dt \tag{B.11}$$

Then, use integration by parts, we have

$$\frac{dLv}{dt} + ad_v^*(Lv) = 0 \tag{B.12}$$

□

References

[1] Peter Bajcsy and Peter Groves. Methodology for hyperspectral band selection. *Photogrammetric Engineering & Remote Sensing*, 70(7):793–802, 2004.

[2] Ruzena Bajcsy and Stane Kovačič. Multiresolution elastic matching. *Computer vision, graphics, and image processing*, 46(1):1–21, 1989.

[3] M Faisal Beg, Michael I Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61(2):139–157, 2005.

[4] J-P Bibring, A Soufflot, M Berthé, Y Langevin, B Gondet, P Drossart, M Bouyé, M Combes, P Puget, A Semery, et al. Omega: Observatoire pour la minéralogie, l'eau, les glaces et l'activité. In *Mars Express: the scientific payload*, volume 1240, pages 37–49, 2004.

[5] Ronald Newbold Bracewell and Ronald N Bracewell. *The Fourier transform and its applications*, volume 31999. McGraw-Hill New York, 1986.

[6] Chaim Broit. Optimal registration of deformed images. 1981.

[7] John Canny. A computational approach to edge detection. In *Readings in computer vision*, pages 184–203. Elsevier, 1987.

[8] Thomas M Cover. Thomas. elements of information theory. *Wiley Series in Telecommunications*, 1991.

[9] Matthias Dantone, Juergen Gall, Gabriele Fanelli, and Luc Van Gool. Real-time facial feature detection using conditional regression forests. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2578–2585. IEEE, 2012.

[10] Paul Dupuis, Ulf Grenander, and Michael I Miller. Variational problems on flows of diffeomorphisms for image matching. *Quarterly of applied mathematics*, pages 587–600, 1998.

[11] Leila MG Fonseca and BS Manjunath. Registration techniques for multisensor remotely sensed imagery. *PE & RS- Photogrammetric Engineering & Remote Sensing*, 62(9):1049–1056, 1996.

[12] Ulf Grenander and Michael I Miller. Computational anatomy: An emerging discipline. *Quarterly of applied mathematics*, 56(4):617–694, 1998.

[13] Peter Groves and Peter Bajcsy. Methodology for hyperspectral band and classification model selection. In *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003*, pages 120–128. IEEE, 2003.

[14] Baofeng Guo, Steve Gunn, Bob Damper, and James Nelson. Adaptive band selection for hyperspectral image fusion using mutual information. In *2005 7th International Conference on Information Fusion*, volume 1, pages 8–pp. IEEE, 2005.

[15] Antoni Rogalski. Infrared detectors: an overview. *Infrared Physics & Technology*, 43(3-5):187–210, 2002.

[16] Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. In *European conference on computer vision*, pages 430–443. Springer, 2006.

[17] John C Russ. *The image processing handbook*. CRC press, 2016.

[18] Alain Trouvé. An infinite dimensional group approach for physics based models in pattern recognition. *preprint*, 1995.

[19] Scott E Umbaugh. *Digital image processing and analysis: human and computer vision applications with CVIPtools*. CRC press, 2010.

[20] Sergio Uras, Federico Girosi, Alessandro Verri, and Vincent Torre. A computational approach to motion perception. *Biological Cybernetics*, 60(2):79–87, 1988.

[21] Barbara Zitova and Jan Flusser. Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000, 2003.

Vita

Rui Liao

Degrees B.E. Beihang University, June 2017

May 2019