Non-rigid registration of mammogram images using Large Displacement Optical Flow with extended flexibility for manual interventions

by

Hae Jin Song

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2018

Author .....
Department of Electrical Engineering and Computer Science
February 3, 2018

Certified by .....
Dr. Regina Barzilay
Professor
Thesis Supervisor

Accepted by .....
Dr. Christopher J. Terman
Chairman, Department Committee on Graduate Theses

Non-rigid registration of mammogram images using Large Displacement Optical Flow with extended flexibility for manual interventions

by

Hae Jin Song

Submitted to the Department of Electrical Engineering and Computer Science
on February 3, 2018, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

This thesis presents a registration method for mammogram images with extended flexibility for manual inputs from medical specialists. The algorithm was developed as part of the Mammography project led by Professor. Regina Barzilay at MIT CSAIL. Given a sequence of mammogram images, the algorithm finds an optimal registration by considering both the global and local constraints as well as user-defined constraints such as manually selected matching points. This allows the registration process to be guided by both the algorithm itself and human experts. The second half of the thesis focuses on evaluating well-known optical flow and medical registration algorithms on mammogram images. It provides insights into how they perform when encountered by challenges and constraints that are unique in mammogram images.

Thesis Supervisor: Dr. Regina Barzilay
Title: Professor

Acknowledgments

I would like to start by thanking Professor.Regina Barzilay for giving me the opportunity of being part of the oncology project and for guiding me through the completion of this thesis. Her passion, dedication and creativity are true motivation to strive for excellence. It has been an honor and joy working in her lab on this project.

I would also like to thank Julian Straub, who introduced me to the group and has helped me from the moment I started, and Erik Nguyen for lending a hand whenever I needed it throughout this journey.

Thanks to Tal Schuster and Adam Yala for brainstorming and helping me for the alignment project.

I am especially grateful for Marcela Rodriguez, Juanjuan Zheng and Manuj Dhariwal for their support, hugs and uncountable great conversations through my time at MIT. It would not have been the same without them.

Lastly, but always first in my heart, thanks to my parents, Yeo Sub Lim and Ha Sook Song, for always being a source of great inspiration and for their unconditional love. This work is for them.

Finally, I would like to thank the Institute for six rewarding years. Getting an education from MIT is indeed like taking a drink from a fire hose.

Contents

List of Figures

Chapter 1

Introduction

Fast and accurate image registration is essential in both medical practices and researches. Its main medical applications include: an early detection of abnormalities, studies of disease progressions by monitoring changes in size and shape over time, image guided surgery or radiotherapy, and development of a statistical model of variation associated with a disease [5]. In this thesis, we focus on the problem of registering mammogram images taken 6 months to a year apart. Our algorithm closely follows the variational framework developed in Brox's Large Displacement Optical Flow [8], and adds extra functionality to intake manual correspondences and regions to be preserved throughout the aligning process. Unlike the original work which was evaluated on the Middlebury dataset [6], we test our modified algorithm on the mammogram images. Even though LDOF reported successful results on estimating motions in the natural images, its success on mammograms bares significance as medical images are different from natural images in the way they are captured, and pose new challenges and constraints. One major difficulty is that the light sources used in medical devices, such as ultrasound or X-ray, go through an obstacle unlike natural lights. The resulting image then becomes a superimposition of multiple objects encountered by light, and there is an added ambiguity as to which objects and to what degree they contribute to the measured intensity value. Natural images, on the other hand, rarely face this problem. Other unique challenges in medical images include the lack of color and shade information, and the long time interval between screenings.

The first portion of this thesis focuses on introducing the problem of non-rigid registration and explaining how we modified LDOF algorithm to allow the registration to be guided by both the algorithm itself and human experts. The second portion focuses on evaluating the modified LDOF and four baseline models (Multi-scale Horn-Schunck, Thirion's Demon, Symmetric Demon and BSpline-based registration) on mammogram images.

Chapter two provides minimal background on non-rigid registration and the variational approach to the problem. Three seminal works are chosen as examples to demonstrate how the variational framework work in image registration.

Chapter three states the mammogram registration problem, as well as its challenges and opportunities. This section directly answers what problem this thesis is tackling and its potential contributions.

Chapter four describes how we modified the original LDOF algorithm [8], and the implementations.

Chapter five demonstrates the registration results from the aforementioned baseline models and our modified LDOF algorithm on mammograms. Quantitative errors from four commonly used metrics are also presented.

Chapter 2

Background

This section provides key background information in image registration with a focus on the non-rigid registration. We first define what is the non-rigid image registration and outline its three main components. Among many approaches to solve this problem, we focus on the variational approach and provide a concise background on the variational calculus and Euler-Lagrange equation. We then present three algorithms (Horn-Schunck [13], Thirion's Demon [21] and BSpline [19]) that are considered foundational in the variational framework. We discuss main assumptions and subsequent limits of the algorithms and follow up with three new concepts that help overcome the shortcomings. We end this section by introducing commonly used evaluation metrics for image registration.

2.1 Non-rigid Image Registration

Image registration is an image processing technique that aligns multiple scenes into a single integrated image [2]. It means to find a mapping between the coordinates in one space and those in another, such that points in the two spaces that correspond to the same 3-dimensional point are mapped to each other. In medical images, an ideal registration establishes the correct mapping between points that correspond to the same anatomical structure. It undoes the effects of different camera perspectives such as scale, rotation and skew, or the change in the physical object's size or po-

sition. Image registration has been studied rigorously in both computer vision and medical imaging communities, and still remains one of the key problems. Its spatial transformation can be categorized into rigid, affine, or non-rigid (or deformable) motion, depending on the space of the parameters it allows. Rigid transformation allows translations and rotations, affine transformation allows skew and scaling in addition to the rigid transformation's parameters, and non-rigid transformation allows free-form mapping. The free-form mapping means that the transformation is not parametrized, and thus each pixel location needs to be assigned a displacement that maps it to the corresponding point on the other image. Naturally, this gives rise to the concept of "displacement field" [1].

Consider an image I as a function of two dimensional pixel location and time, $I(x, y, t)$. I_1 and I_2 are the two images that are taken at time t and $t + 1$ so that:

$$I_1(x, y) = I(x, y, t)$$
$$I_2(x, y) = I(x, y, t + 1)$$

Let I_1 be the target (or the fixed) image and I_2 the source (or the moving) image. A displacement field $\vec{w}(x, y) = [u, v]^T$ is a function on I_1 that maps each location in I_1 to its corresponding location in I_2:

$$I_1(x, y, t) = I_2(x + u, y + v, t + 1), \quad \forall x, y \in \Omega$$

where Ω is the image domain, and u and v are functions of x, y: $u = u(x, y)$, $v = v(x, y)$.

Image registration is often referred to as optical flow estimation in computer vision. We will use the two terms interchangeable throughout the thesis.

[1] We use displacement field, velocity field and flow field interchangeably

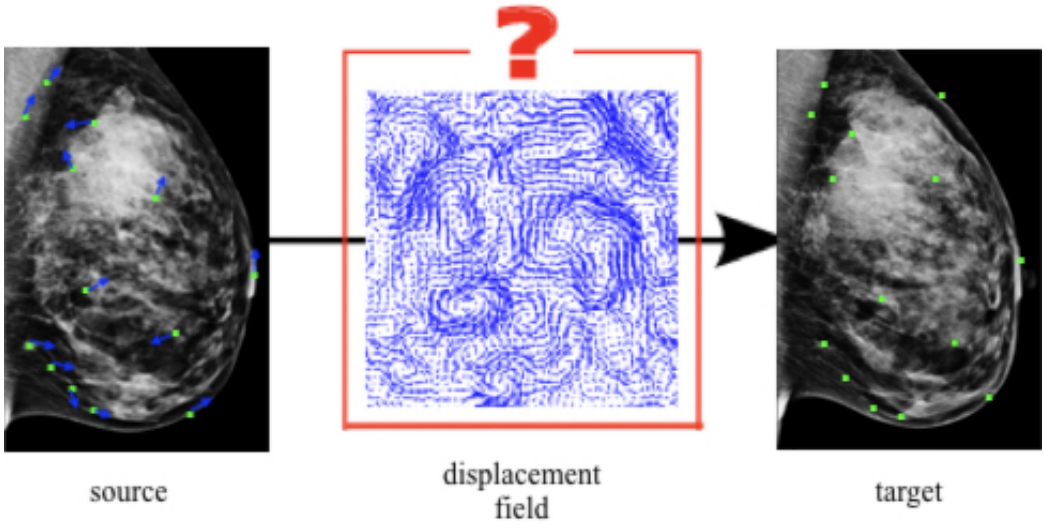


Figure 2-1: Image registration is a spatial transformation that maps points from one image to corresponding points in another. Non-rigid registration defines a free-form mapping between the source (or "moving") and the target (or "fixed") images via a displacement field.

2.1.1 Variational approach

Among many approaches to solve the image registration problem, we focus on the variational approach, which defines an energy function between the source and the target and views the registration as an optimization of the energy function. Horn and Schunck first proposed this approach to optical flow estimation in 1981 [13], and it still remains the choice of most advanced optical flow algorithms. It has an advantage that the objective function is well-defined and thus the solution is interpretable. In addition, a plethora of optimization methods become available as a tool to solve the optimization problem. The variational framework can be broadly decomposed into the following three components: deformation model, matching criteria and optimization method [9].

A deformation model specifies the way in which the source image can be changed to match the target. In non-rigid registration, physic-based models (such as elastic, fluid and optical flow) and models using basis functions or splines exist [5].

The matching criteria measures how similar two images are and is used to define

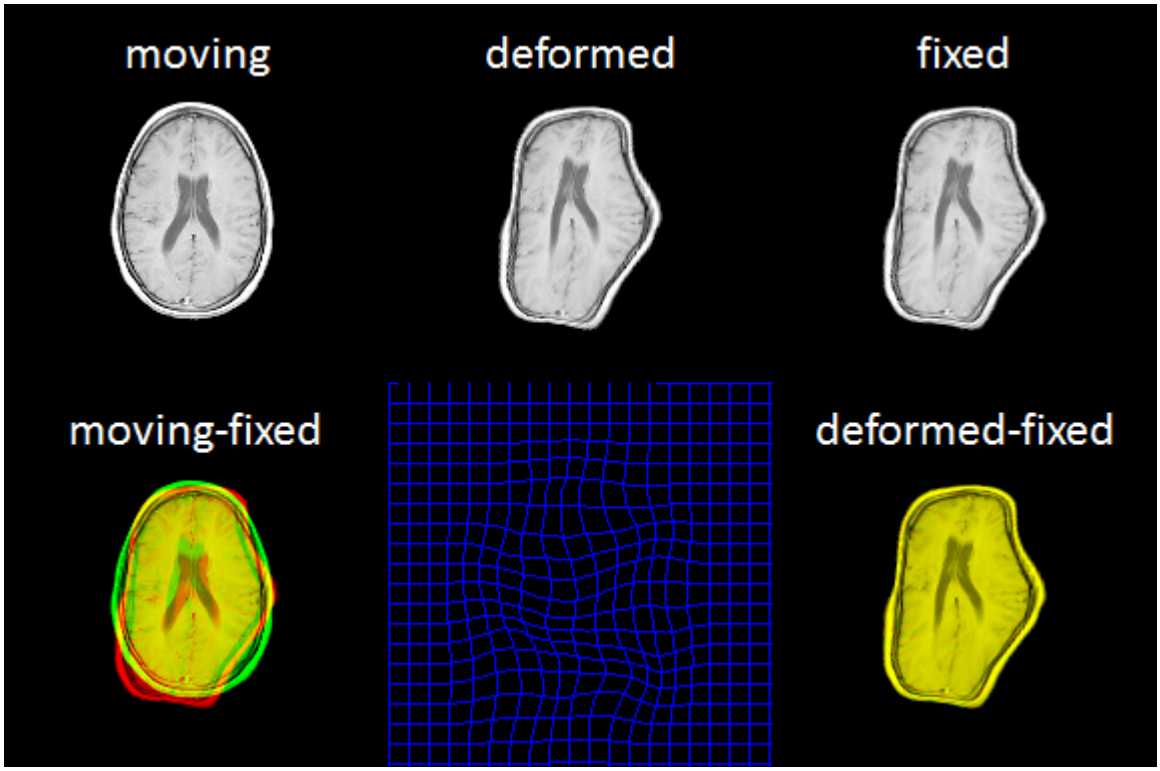


Figure 2-2: Non-rigid registration demonstration. The registration process finds a displacement field that maps each location in the moving image I_2 to the corresponding location in the fixed image I_1. Ideally, the location pair corresponds to the same anatomical structure [3]

the objective function for the optimization process. We may categorize the metrics into the following four categories. First, landmark-based metrics compute the distance between the sets of keypoints extracted from each image. Surface-based metrics extract corresponding surfaces between the two images and use their distances as a similarity measure. Intensity-based metrics are computed directly from the pixel values in the images, and commonly used metrics such as Sum of Squared Difference (SSD), Sum of Absolute Difference (SAD), Normalized Cross-Correlation (NCC) and mutual information (MI) all belong to this category. Feature-based metrics are based on the correspondence between image features such as points, lines, and contours. Similarity metric between feature values such as curvature can be used.

The optimization process varies the parameters of the deformation model to maximizing the matching criterion. Many registration algorithms require an iterative approach to reach an optimal solution. The process starts with an initial estimate for the displacement field $\vec{w_0}$, and gradually refine the estimate. At each iteration, the current estimate is used to calculate a similar the target and the current deformed model are. Next estimate of $\vec{w}$ is made based on this similarity measure until the algorithm converges or reaches a termination condition [5]. Optimization methods vary in their choice of the update step for the next w. The algorithms discussed in the thesis are based on gradient-descent methods such as the Steepest Descent, Regular step gradient descent (RSGD) [22], and quasi-Newton method (LBFGS [16]). More in-depth discussion on the non-rigid registration can be found in [9]. Now, we present two seminal works that are formulated in the variational framework and highlight each of their three components. The first algorithm by Horn-Schunk is the first variational method for optical flow estimation, and the second algorithm by Thirion is one of the pioneering work in deformable registration of medical images.

2.2 Horn-Schunck

Horn-Schunck's optical flow [?] views the optical flow as the solution of a minimization problem. Its main assumption, often referred to as "brightness consistency" assumption, is that pixels intensities do not change over time. Therefore, the corresponding points in I_1 and I_2 will have the same intensity value. Using the first order Taylor expansion for the intensity function, Horn-Schunck derives the *optical flow constraint* from this assumption:

$$\begin{aligned} I(x,y,t) &= I(x+\Delta x, y+\Delta y, t+\Delta t) \\ &= I(x,y,t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t + \epsilon \end{aligned}$$

where the linerization using the first-order Taylor expansion occurs in the second line. Subtracting $I(x,y,t)$ from both sides and dividing by Δt gives,

$$\begin{aligned} 0 &= \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t \\ &= \frac{\partial I}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial I}{\partial y}\frac{\Delta y}{\Delta t} + \frac{\partial I}{\partial t}\frac{\Delta t}{\Delta t} \end{aligned}$$

As $\Delta t \to 0$,

$$0 = \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} \tag{2.1}$$

where u and v can be the velocity of the motion in x and y direction, respectively.

However, the optical flow constraint leads to an ill-posed problem as we have two unknowns (u, v) at each pixel location with a single constraint 2.1. To overcome this problem and achieve a unique solution, Horn-Schunk adds an additional regularity term that measures the smoothness of the flow field. The smoothness is measured by the magnitude of the flow's spatial gradient. The final energy function thus is defined as follows:

$$\int_\Omega (I_x u + I_y v + I_t)^2 + \alpha^2(\|\nabla u\|^2 + \|\nabla v\|^2) \tag{2.2}$$

where α is the parameter that controls the relative weight of the brightness consistency

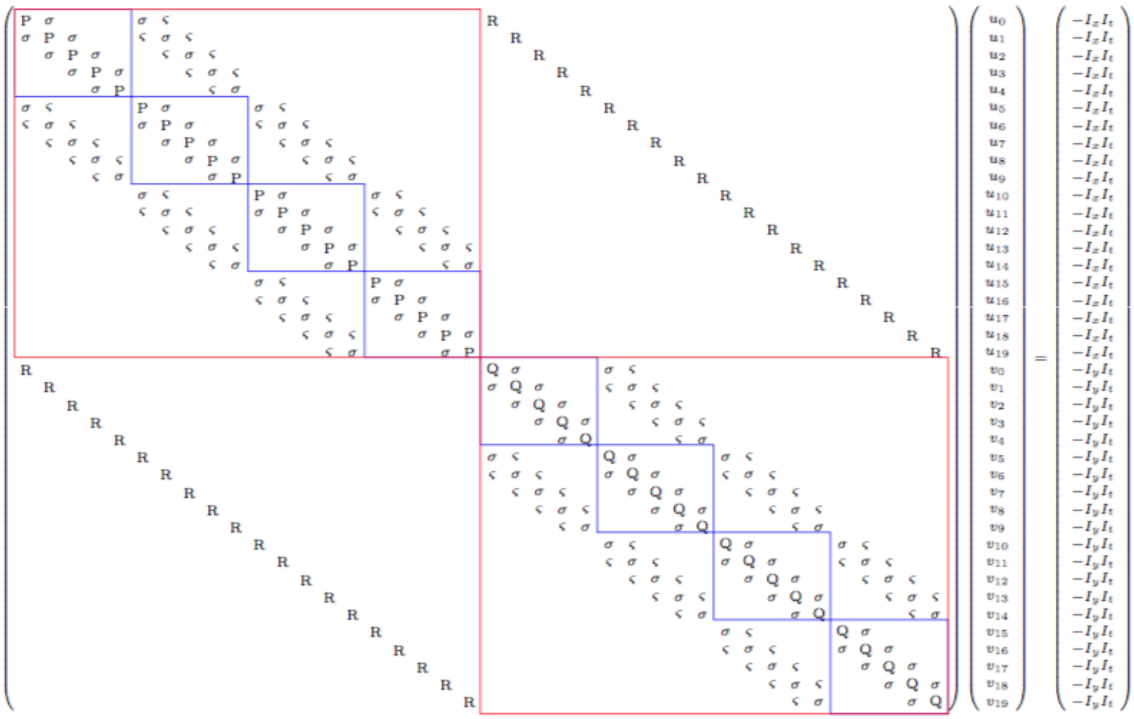


Figure 2-3: Linear system for Horn-Schunck optical flow. Since each pixel has two equations to solve, this *sparse* matrix is of size $2n$ by $2n$ where n is the number of pixels. An iterative method such as Gauss-Seidel or Jacobi method is used in practice. Reprinted from [17]

condition and the smoothness.

The minimization of the above functional results in the following Euler-Lagrange equations:

$$I_x^2 u + I_x I_y v = \alpha^2 \text{div}(\nabla u) - I_x I_t \tag{2.3}$$

$$I_x I_y u + I_y^2 v = \alpha^2 \text{div}(\nabla v) - I_y I_t \tag{2.4}$$

By solving Equation 2.3 for each $(x, y) \in \Omega$, we achieve a dense optical flow over the entire image domain. It amounts to solving the following $2n$ by $2n$ (where n is the number of pixels) linear system of equations:

The size of this matrix is however too large to store and directly solve in practice. For instance, if an image is of size 128 by 128 ($n = 128^2 = 16384$), the size of the matrix becomes larger than 1 billion: $2n \times 2n = 1,073,741,824$. Nevertheless since

the matrix is sparse, the solution can be iteratively achieved via Gauss-Seidel or Jacobi methods.

The main advantages of Horn-Schunck optical flow is that it gives a dense flow estimation even in the area where the gradient is zero such as textureless or homogeneous surfaces. On the other hand, it is sensitive to noise and outliers due to the choice of L^2 norm to measure the intensity difference. The two assumptions mentioned above also make a compromise with the algorithm's capability. Because the brightness consistency assumption depends on the exact intensity value (rather than a relative one), it cannot intelligently compensate a global change in the lightening. The linearization by the first-order Taylor expansion makes the estimation exclusively valid for a "small"(< 1 pixel) motion. In later section (Section **??**), we will introduce some methods that can handle these limitations.

2.3 Demon's method

Demons method was first introduced by Thirion in 1998 [21] and extended by Cachier 1999 and He Wang 2005, to name a few. Demons-based methods view the image registration as a diffusion process during which the object boundaries in the fixed image acts as semi-permeable membranes through which the moving image diffuse. Each pixel in the fixed image act as local forces (as if applied by "demons" in Maxwell's equation [21]) that can move the pixels in the source to match the intensities in the target.

This displacement field is smoothed by an Gaussian filter at each step. Thirion used the following update step to compute the incremental displacement field:

$$d\vec{w}^{(n+1)} = \frac{I_2^{(n)} - I_1^{(0)}}{(I_2^{(n)} - I_1^{(0)})^2 + \|\nabla I_s^{(0)}\|^2} \vec{\nabla} I_s^{(0)} \tag{2.5}$$

As in Horn-Schunck's optical flow equation, Equation 2.5 is under-determined. To find a unique solution, Thirion proposed to filter the displacement field at each iteration with a Gaussian filter.

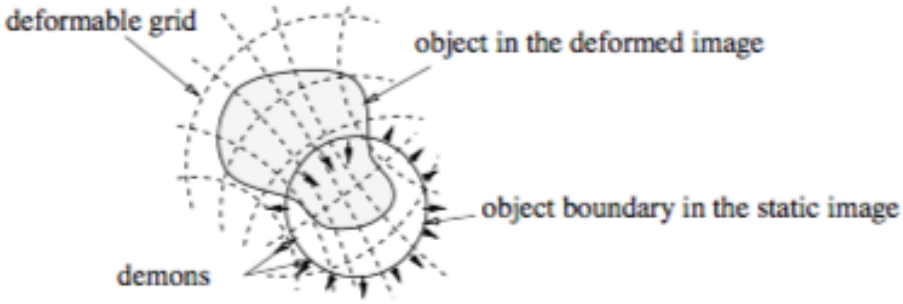


Figure 2-4: Diffusion model in Demon's method. The moving image diffuses through the boundaries of the objects in the fixed image, controlled by the forces of effectors called "demons". Reprinted from [21]

Better theoretical understandings on the demon's method came from Pennec's work in the following year [18]. In this work, Pennec showed that the forces proposed by Thirion correspond to a second order gradient descent on the sum of square of intensity differences (SSD) criterion. He also discussed that the Gaussian smoothing acts as a greedy optimization of the regularized criterion.

Note the displacement field is derived from the source image alone in Thirion's original demons. Many variants of demons have been developed by modifying how to compute this force. One of the baseline models used in Section 5.1 is Symmetric Demons by Vercauteren [23] which uses both the source and the target images to derive the deformation force.

2.4 BSpline

A BSpline-based method is another approach to non-rigid registration based on the deformations on the control points and bpline interpolation. First, a grid of control points is constructed at the selection resolution which controls the transformation of the source image. The similarity metric is used to measure the registration error between the current moving and static image, and the quasi newton optimizer such as LBFGS [16] is used to move the control points to achieve the optimal registration with minimal registration error. B-spline registration is slower and more complex than Thirion's Demond. However, it is still used because the resulting displacement field

corresponds to a better real-live deformation that the one from fluid-based methods like Demon's algorithm [15].

Chapter 3

Mammogram registration problem

This thesis tackles non-rigid registration of medical images by focusing on the mammogram images taken certain time intervals apart. Given a sequence of mammogram images of either 1) different patients ("intrapatient") or 2) the same patient over time ("interpatient"), we aim to align the images without losing key structures and important information for medical diagnosis. In addition, we evaluate both traditional and recent registration algorithms on mammogram images and compare their strengths and weaknesses. The five registration methods we compare are 1) Multiscale Horn-Schunck (MSHS), 2) Demon's method [21], 3) Symmetric Demon's method [23], 4) BSpline [19, 20], and 5) Large Displacement Optical Flow (LDOF) [8].

As discussed in Section **??**, the problem of image registration has been studied by both computer science and medical imaging communities over decades. More advanced techniques have complemented drawbacks of traditional approaches such as Horn-Schunck's optical flow [13], yet registering medical images is still an unsolved problem. One of the major difficulties lie in the validation. Optical flow is inherently an ill-posed problem [**?**], which means there are many deformation fields that result in reasonable registrations. However, there is no established way to objectively evaluate these fields and choose the "right" one for a particular application [9]. Currently, most algorithms are evaluated on the standard benchmark image data sets such as Middlebury data [6] and KITTI data [10], which are limited to natural scenes and do not include any medical images. Consequently, their performances on medical images

hasn't been rigorously tested. Therefore, this thesis consists of two parts. First, it extends Large Displacement Optical Flow [8] – an algorithm that incorporates automatic feature extraction and descriptor matching to the conventional variational framework – so that it considers manual inputs from medical specialists during the registration process. The second part focuses on evaluating the four algorithms mentioned above on mammogram images. They are chosen based on their significance and popularity in computer vision and medical imaging applications.

Surprisingly, only a limited amount of optical flow algorithms have been used in the medical applications. The first portion of this thesis reports the success of Large Displacement Optical Flow on mammogram registration, and discusses how we extend it to also consider direct inputs from medical specialists to adjust the registration. The success of the LDOF on mammogram registration is not an obvious result, even with its previous success with natural images, because medical images impose constraints and requirements that are different from natural images. One major difference is that light sources such as ultrasound or X-ray go through an obstacle unlike natural lights. Thus a medical image is a superimposition of multiple objects encountered by the light as it travels, and there is an added ambiguity on which objects and to which degree they contribute to a pixel value. Natural images, on the other hand, rarely face this problem. More efforts are put into accurate estimations on the boundary where occlusions occur. Another unique challenge in medical images is the lack of color and shade information. Color and shading from natural images have been shown beneficial in estimating shapes and motions [12],[11]. Most of the current medical devices are not able to capture those information.

Fast and accurate image registration is essential in both medical practices and researches. Its main medical applications are: an early detection of abnormalities, studies of disease progressions by monitoring changes in size and shape over time, image guided surgery or radiotherapy, and development of a statistical model of variation associated with a disease [5]. The success of the original LDOF algorithm in registering mammograms, which will be represented in Section 5.2, means a fast, annotation-free registration. Our extension makes the original algorithm more usable

and controllable by the radiologists. It gives the flexibility for them to manually specify corresponding pairs, or regions that need to be preserved during the registration. This modification allows the registration to be guided by both the algorithm itself and the human experts.

The second focus of the thesis bares its own significance in that it make a head-to-head comparison among traditional optical flow algorithms and medical alignment methods on the same type of images (i.e mammograms). Due to the difference in their focus and applications, these methods are often not evaluated on the same type of images. Therefore, we gain new insights on their strengths and weaknesses when applied to medical images. Lastly, the evaluation section demonstrates how the commonly used metrics contradict each other and do not agree with human evaluation. This brings attention to the outstanding problem of validation in non-rigid image registration.

Chapter 4

Methodology

4.1 Extented LDOF: Manual correspondences and region preservation

Let $I_1, I_2 \in \mathbb{R}_2$ be the two images to be aligned. The first image is fixed (ie. target), and we aim to register the second image (source) to the first image. An optimal dense displacement field is the one that deforms the second as "close" to the first image as possible. Let $\vec{x} = [x, y]^T$ denote a pixel location and $\vec{w} = [u, v]^T$ a displacement field in the image domain Ω. Note $\vec{w}$ is a function of x and y. First we assume that a pixel's intensity does not change at the corresponding locations in I_1 and I_2. To compensate any global illumination change, we also assume the magnitude of (local) change in the color remains constant at corresponding pixel locations. These two assumptions are incorporated into the energy function as follows:

$$E_{data} = \int\int_{\Omega} p(I_1(\vec{x}) - I_2(\vec{x} + \vec{w})) + \alpha q(\nabla I_1(\vec{x}) - \nabla I_2(\vec{x} + \vec{w}))d\vec{x} \qquad (4.1)$$

We choose $p(x)$ and $q(x)$ to be $\phi(x^2) = \sqrt{x^2 + \epsilon^2}$ with $\epsilon = 0.001$ as in [8] and [7]. In Horn-Schunk optical flow [13], $p(x) = x^2$, and many other algorithms that uses TV-L1 as a similarity metric [24] use $p(x) = (x)$. Our choice of p, q allows us to deal with occlusions and non-Gaussian deviations of the similarity criterion [8]. The

extra ϵ term is fixed to 0.001 to maintain the energy function convex while voiding potential division by zero failures in areas, for instance, with constant gradients or few structural patterns. This choice of ϕ leads to a L^1 norm minimization and convex energy function [7]. We would also like to emphasize that the energy function defined in 4.1 is non-linear with respect to $\vec{w}$ due to its dependency on the intensity functions. Unlike Horn-Schunck algorithm which use the first-order Taylor approximation to linearize the equation as discussed in Section 2.2 we embrace the non-linearity and allows the algorithm to handle both small and large deformations. The parameter α controls the weight on the gradient consistency constraint.

However, minimizing the difference measure in 4.1 with respect to $\vec{w}$ is an ill-posed problem, because at each pixel we have two unknowns $\vec{w} = [u, v]^T$ but only a single equation. In order to find a unique solution, we add a regularization term. Specifically, we let the algorithm to prefer a globally smooth motion field by adding the energy term that measures the deviation from a constant field:

$$E_{smooth}(\vec{w}) = \int\int_{\Omega} \phi(\|\vec{\nabla}u\|^2 + \|\vec{\nabla}v\|^2)d\vec{x} \tag{4.2}$$

Note that $u = u(x, y)$ and $v = v(x, y)$.

Beyond that the regularization is necessary to find a unique solution $\vec{w}$, it allows us to incorporate prior knowledge, for instance about underlying tissue properties during deformation. Although we did not take advantage of such capability due to the lack of an established model for breast tissue, future work can be extended to incorporate relevant biological models. It also reduces the number of local minimum [5] and makes the optimization methods better converge to the global minimum.

As in the original LDOF algorithm in [8], we incorporate the problem of correspondence establishing and descriptor matching into the variational framework. First we define the descriptor matching term E_{desc}:

$$E_{desc}(\vec{w}) = \int\int_{\Omega} \delta_{desc}(\vec{x})p_{desc}(\vec{x})\phi(\|\vec{w}(\vec{x}) - \vec{w}_1(\vec{x})\|^2)d\vec{x} \tag{4.3}$$

$\delta\vec{x}$ is the descriptor indicator function whose value if 1 if $\vec{x}$ is a feature point and 0

otherwise. $\vec{w_1}(\vec{x})$ is the difference vector obtained by matching $\vec{x}$ in I_1 to its corresponding pair in I_2, and $p(x)$ is the weight computed for the descriptor at $\vec{x}$ defined as in [8]. This energy term assumes that 1) features (and descriptors) are extracted from both I_1 and I_2, and 2) correspondences are established between the two sets of features. The second part, of establishing the correspondences, is formulated as another energy term E_{corr}:

$$E_{corr}(\vec{(}w), \vec{w_1}) = \int\int_{\Omega} \delta_{desc} \|(\vec{f_2}(\vec{x} + \vec{w_1}(\vec{x})) - \vec{f_1}(\vec{x}))\|^2 d\vec{x} \tag{4.4}$$

where $\vec{f_1}(\vec{x})$ and $\vec{f_2}(\vec{x})$ are sparse fields of feature vectors in I_1 and I_2. As an extension to the LDOF, we collect two corresponding keypoints from I_1 and I_2 and define the boundary of regions that should be preserved during the transformation. Using these user inputs, we formulate the user-defined keypoint matching and region preservation constraints as the two following terms, E_{user} and E_{region}:

$$E_{user}(\vec{w}) = \int\int_{\Omega} \delta_{user} q_{user}(\vec{x} p(\|(\vec{w}(\vec{x}) - \vec{k}(\vec{x})\|^2) d\vec{x} \tag{4.5}$$

where δ_{user} is an indicator function for manually selected keypoints, and q_{user} is a weight function on each selected keypoint pair, corresponding to their significance or confidence level.

$$E_{box}(\vec{w} = \int\int_{\Omega} \delta_{box} q_{box}(\vec{x}) \|\vec{w}(\vec{(}x)\|^2 d\vec{x} \tag{4.6}$$

where δ_{region} is an indicator function, and q_{region} is the weight mask for the pixels in the region. We used Gaussian distribution with σ of 3 in our experiments. Putting

together all energy terms gives the following final energy function:

$$E(\vec{w}) = E_{color}(\vec{w}) + \alpha E_{grad}(\vec{w}) \tag{4.7}$$

$$= \gamma E_{smooth}(\vec{w}) \tag{4.8}$$

$$= \beta_1 E_{corr}(\vec{w}, \vec{w_1}) + E_{desc}(\vec{w_1}) \tag{4.9}$$

$$= \beta_2 E_{user}(\vec{w}, \vec{k}) \tag{4.10}$$

$$= \beta_3 E_{box}(\vec{w}, B) \tag{4.11}$$

where $\vec{k}$ and B are fixed inputs from the user, the manual keypoint correspondences and the selected bounding box.

4.2 Optimization

Since the two added energy terms are convex, the entire energy function remains convex. As in the original paper [8], we use the continuation method to solve this optimization problem. Specifically, we use the downsampling factor of $0.95^{(k_{max}-k)}$ to allow a very smooth transitions between consecutive pyramid levels, where k_{max} is chosen for valid discrete derivative filters. We then derive a Euler-Lagrange equation for u, v and use the nested fixed iterations to remove the nonlinearity from the intensity functions.

Chapter 5

Evaluation

5.1 Baseline models

In this section, we compare the registration result from LDOF algorithm [8] with five baseline models. We compute the registration scores using four widely used metrics to compare their performances on mammogram images. Our baseline models are 1) Multi-scale Horn-Schunk (MSHS), 2) Demon's method with four levels of resolution, 3) Symmetric demons with four levels of resolution, 4) 3 level BSpline with LBFGS optimizer, and 5) 3 level BSpline with regular step gradient descent (RSGD) optimizer. The demons- and BSpline-based baseline models were tested using ITK [14] after minor modifications on the implementations in [1].

The result from Multi-scale Horn-Schunk (MSHS) in Figure 5-2a demonstrates undesirable holes, and fails to preserve important features of the images. For medical purposes, the key feature is the tumor, which is marked with the red bounding box in the orignal I_2 (Figure 5-1. MSHS's result removes this distinctive feature during registration since its objective function is defined globally. The output image resembles a simple superposition of the I_2 onto the original I_1. Another limit of MSHS is its lack of awareness on the anatomical structure. Since its objective function is defined only with respect to intensity consistency and the smoothness of the deformation field without any notion of partial structures, the result is not guaranteed to establish reliable correspondences between anatomical structures.

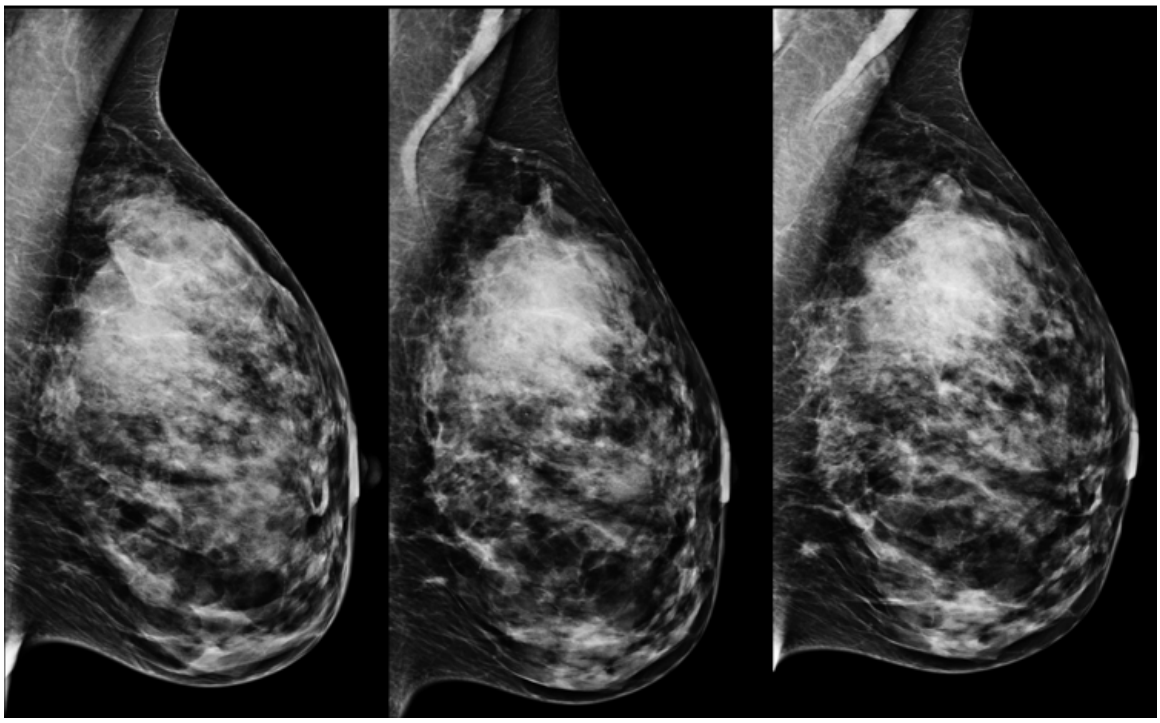
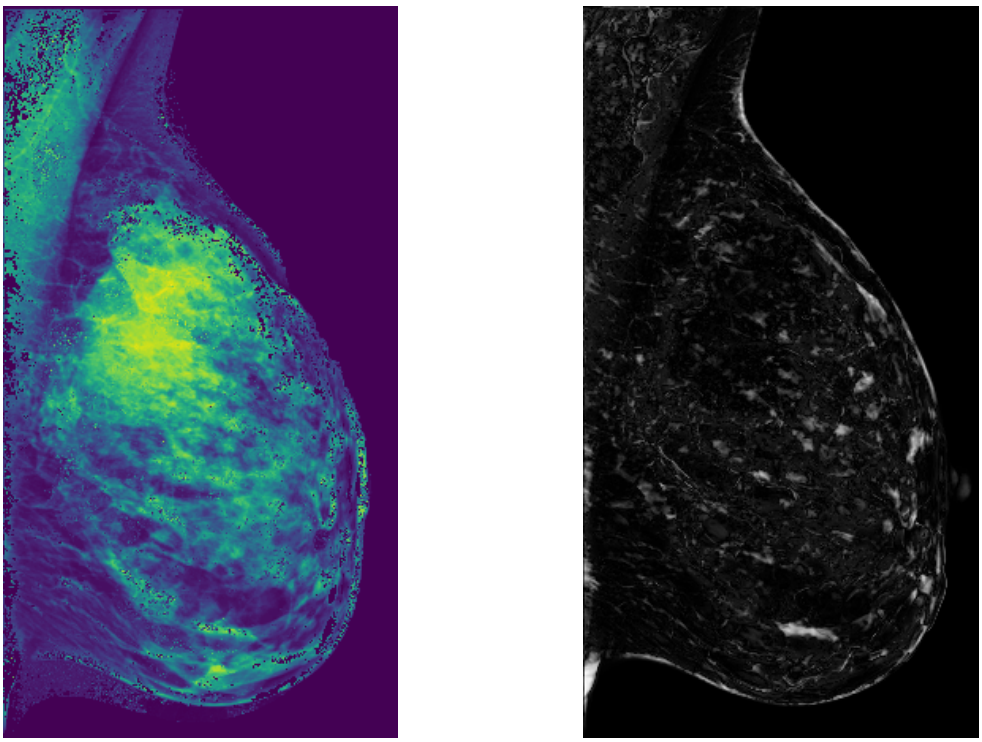


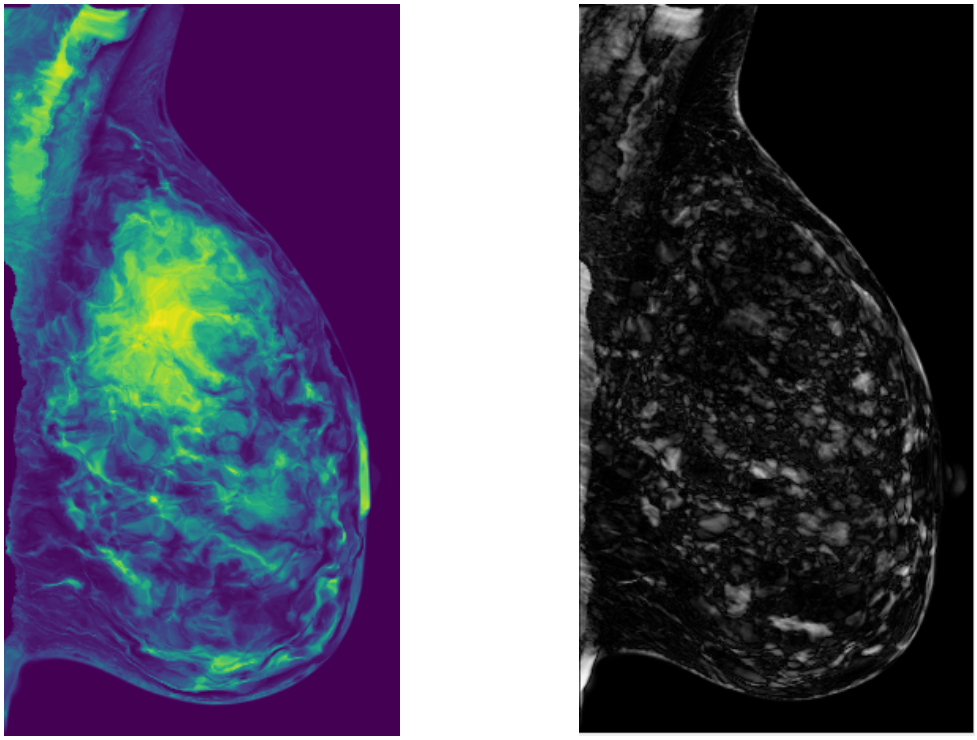
Figure 5-1: Original mammogram sequences

The result from Demon's algorithm in Figure 5-3a does not show the same problem of holes during the backward warping process as observed in Horn-Schunk. However, the structure on the left end is lost, and the white tissues diffuse to the entire image. This diffusion of the white tissues is the result of Thirion's view on the non-rigid registration as a diffusion process [21]. The diffusion model is however not shown to truthfully reflect the anatomical displacements during the mammogram imaging. In comparison, the Symmetric Demons method provides a result with better defined tissue clusters as shown in Figure 5-4a. However, the algorithm has the same limit of relying on the intensity alone (i.e. minimizing the SSD criterion via the second order gradient descent and adding regularization [18]). As discussed in Section 4, LDOF overcomes this drawback by explicitly defining the feature matching and descriptor preservation scores and adding it to the objective function. More discussion on the results of LDOF follows in Section 5.2. Figure 5-5a shows the result of multi-scale BSpline registration with LBFGS optimization. The result shows a large warping on the background which implies that the algorithm or the optimization process were



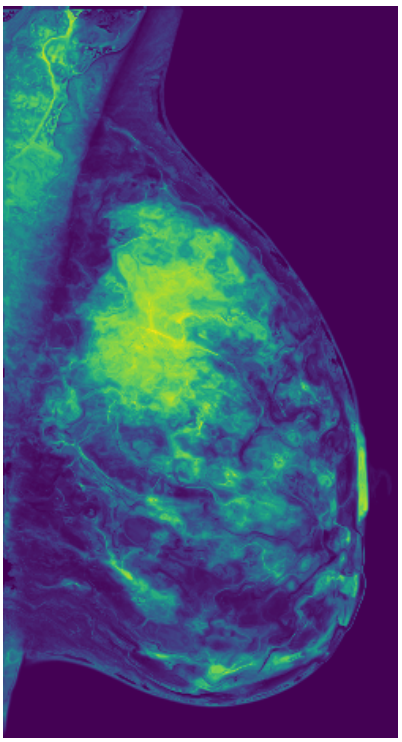
(a) warped I_2 (b) diff

Figure 5-2: Multi-resolution Horn-Schunck

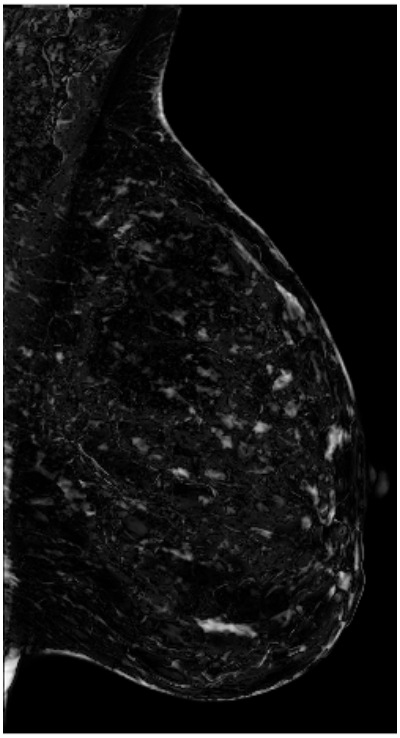


(a) warped I_2 (b) diff

Figure 5-3: Demons

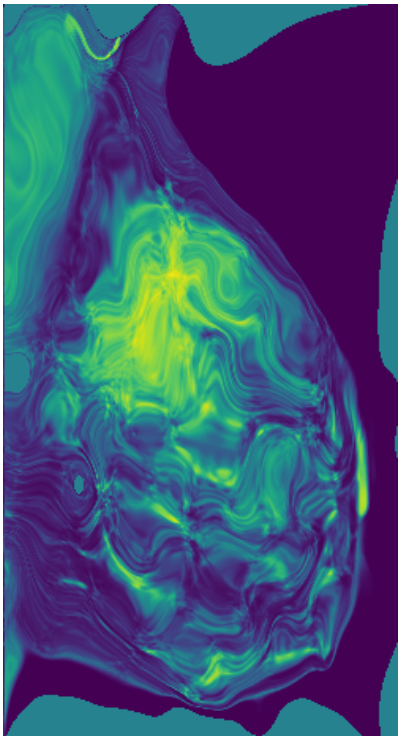


(a) warped I_2

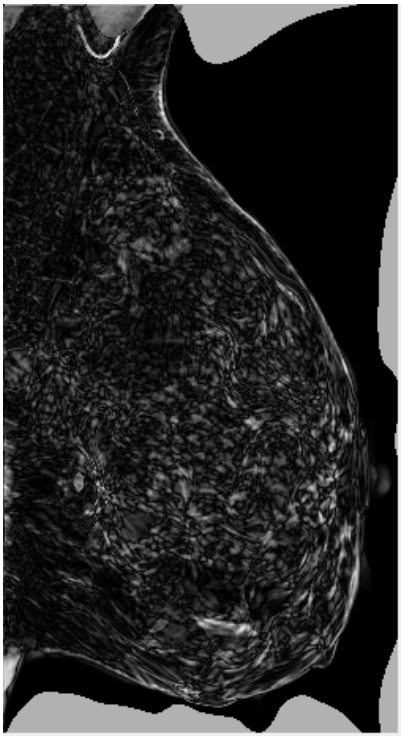


(b) diff

Figure 5-4: Symmetric Demons

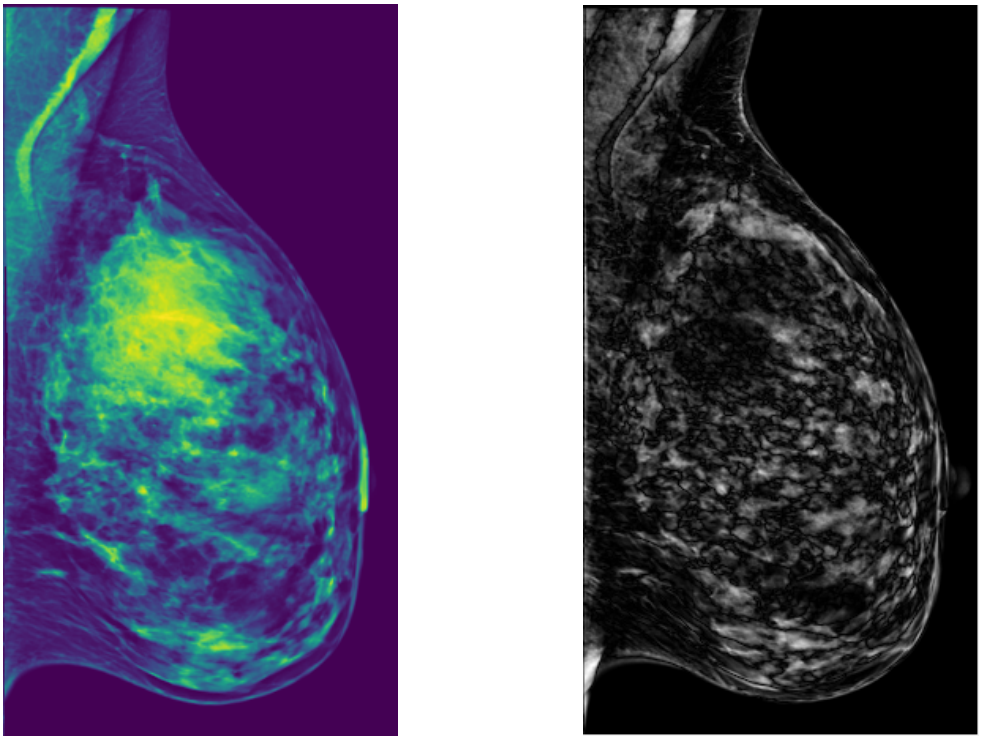


(a) warped I_2



(b) diff

Figure 5-5: BSpline with LBFGS



(a) warped I_2 (b) diff

Figure 5-6: BSpline with Regular Step Gradient Descent

strongely affected by noise. It is however worthwhile to note that the tumor region is well-preserved.

The next figure (Figure 5-6a shows the result of the same multi-scale BSpline registration but with regular step gradient descent optimizer. The result doesn't suffer from the incorrect deformations in the background which was prevalent in the BSpline registration with LBFGS optimizer. Yet, a closer observation in Figure 5-7 reveals that the tumor regions are misaligned.

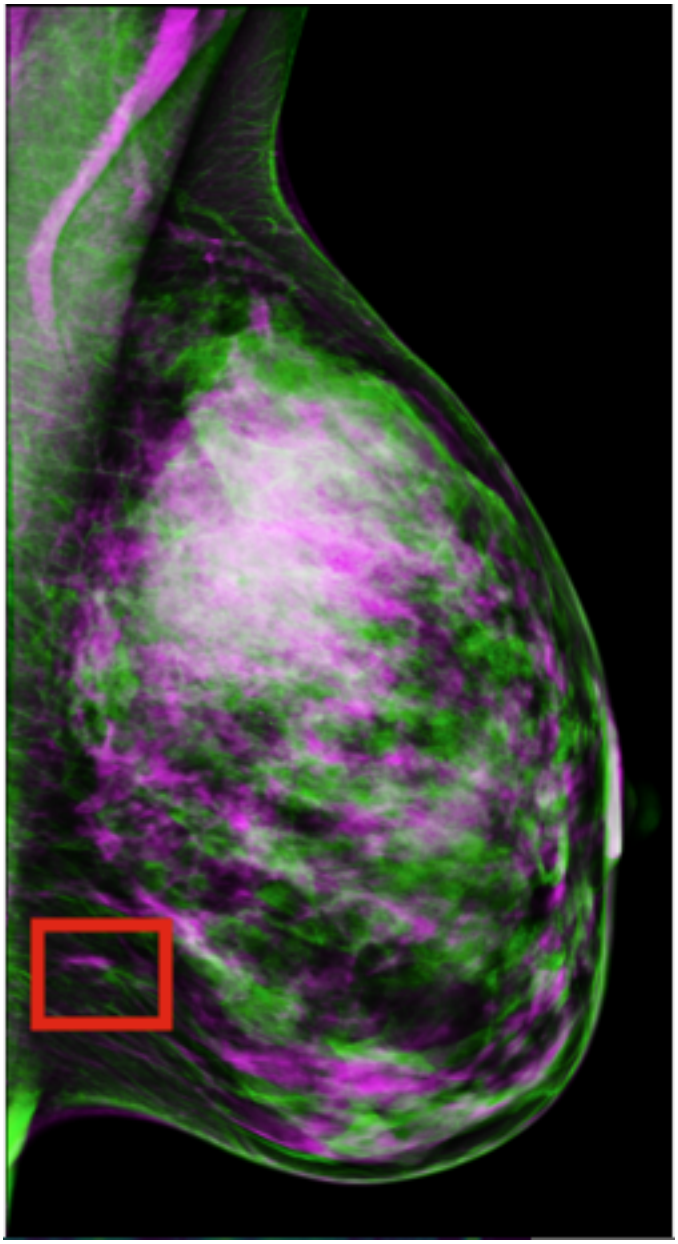
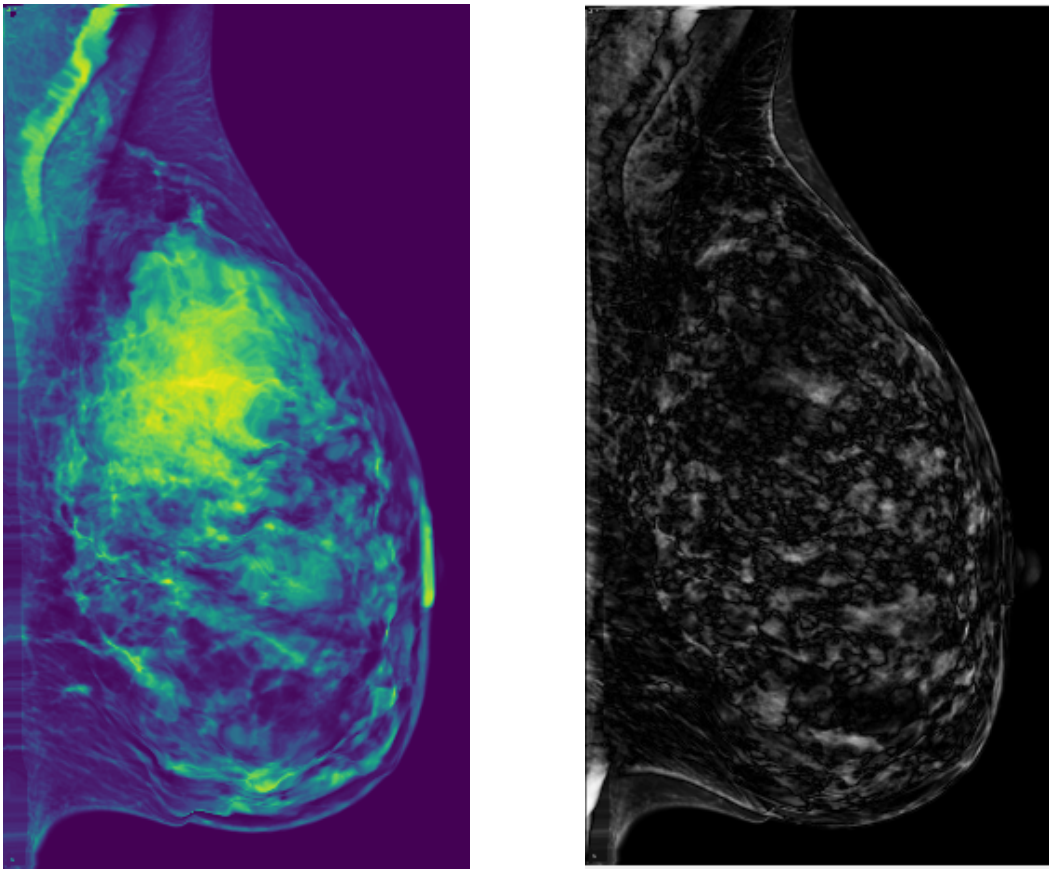


Figure 5-7: I_1 and registered image I_2 using BSpline with Regular Step Gradient Descent. The tumor marks in the red bounding box are misaligned.

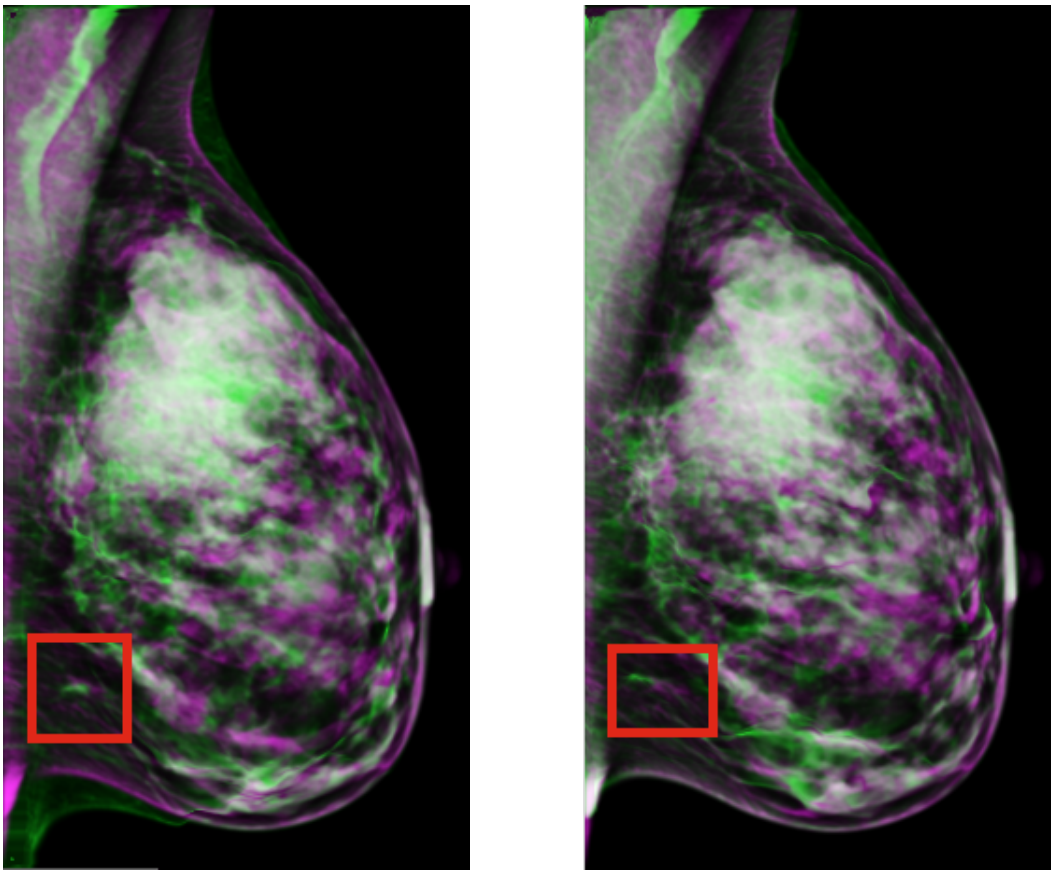
5.2 Registration using LDOF



(a) warped I_2 (b) diff

Figure 5-8: LDOF

In comparison to the baseline models' registrations, the result from LDOF achieves both global and local alignments. Figure 5-9 shows the superposition of the original I_1 and registered images of I_2 and I_3, respectively. The overall shape of the breasts and the static structures on the left end are well-aligned. At the same time, the local correspondences between matching key points are also preserved as highlighted in the red bounding box. The tumor regions inside the boxes overlap with much smaller displacement than what is achieved by baseline models. In particular, it finds a better alignment than Symmetric Demon's algorithm (Figure 5-7) which achieved the best result among the baseline models. Compared to Demon's algorithm (Figure 5-3a) which leads to the diffusion of the white tissues driven by the intensity difference,



(a) I_1 and warped I_2 (b) I_1 and warped I_3

Figure 5-9: The tumor marks in the red bounding boxes are well-aligned as well as the overall shape of the breast. See Figure 5-7 to compare with the result from BSpline registration.

LDOF restricts the degree of diffusion and optimize for more accurate feature matching and more conservative descriptor preservation.

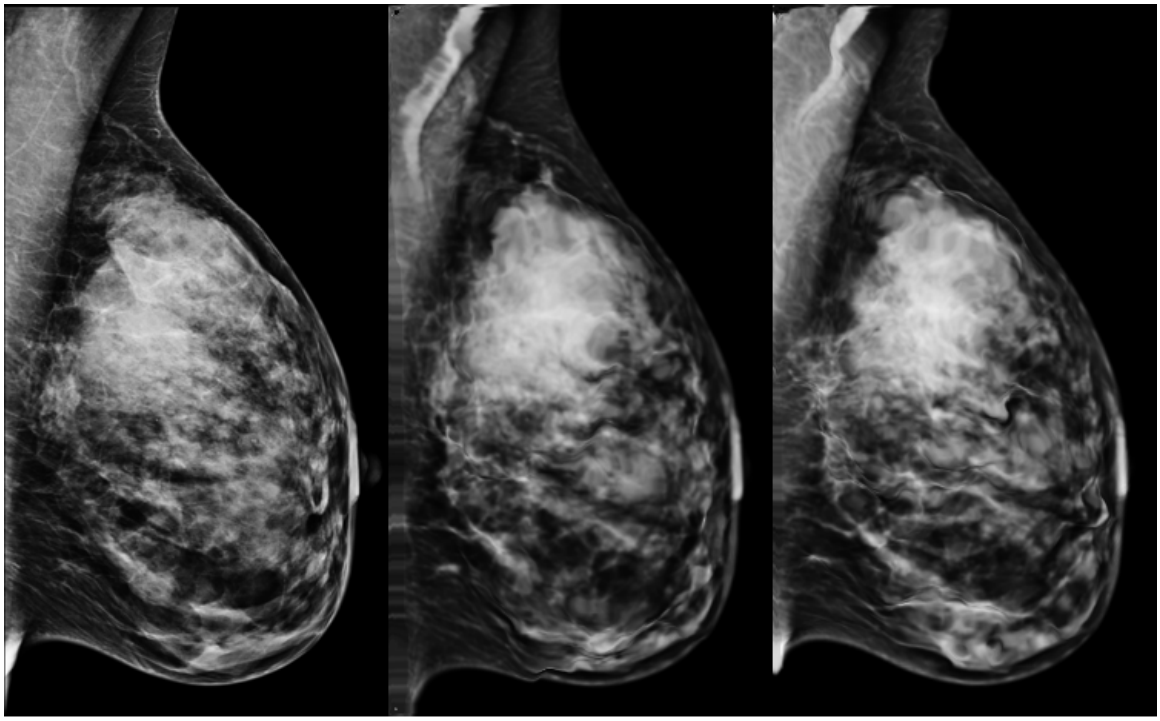


Figure 5-10: LDOF: Registered sequence

5.3 Quantitative evaluation

This section provides quantitative evaluations on the baseline models and LDOF registration. In Section 2, we discussed how each registration method optimizes a different objective function. The goal of this section is to show how evaluation of the registered image based on a single metric fails to correctly measure the quality, and highlight the need for a more unified metric that is independent of the choice of metric used in the registration process itself, which will allow to evaluate various registration methods on the common basis. We evaluate the six registration methods using four popular metrics: sum of squared (intensity) difference (SSD), sum of absolute difference (SAD), normalized cross-correlation (Corr), and mutual information (MI). SSD and SAD are computed on float images after scaling them to $[0, 1]$:

$$SSD = \frac{1}{N} \sum_{x,y} (I_1(x, y) - \text{warped}\mathrm{I}_2(x, y))^2$$

$$SAD = \frac{1}{N} \sum_{x,y} |I_1(x, y) - \text{warped}\mathrm{I}_2(x, y)|$$

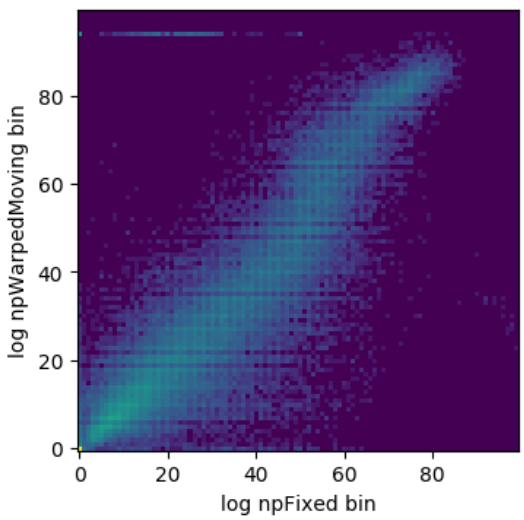
Information theory based approach to image registration (as opposed to feature-based or intensity-based, see Section 2) tries to maximize the amount of shared information in two images. Mutual information (MI) is one of the metrics that measure the amount of shared information between two signals (or random variables in general), or how well one image explains the other. Information theory based approach postulates this measure is maximized at the optimal alignment. It is defined as:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)\, p(y)} \right)$$

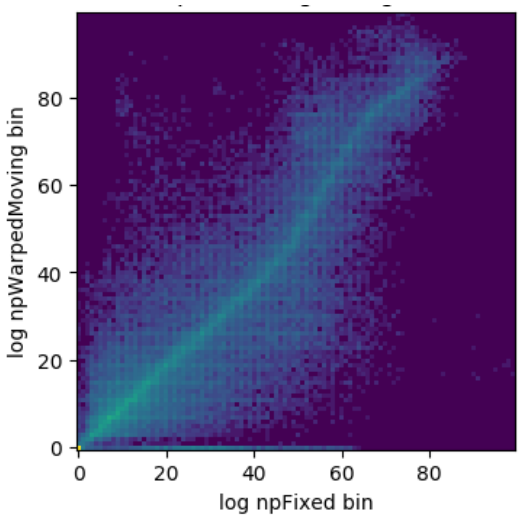
The metric is high when the signal is highly concentrated in few bins, and low when the signal is spread across many bins. In practice, it is computed from the joint 2D histogram of the images in the following way [4]:

```
def mutual_information(hgram):
    # Mutual information for joint histogram
    # Convert bins counts to probability values
    pxy = hgram / float(np.sum(hgram))
    px = np.sum(pxy, axis=1) # marginal for x over y
    py = np.sum(pxy, axis=0) # marginal for y over x
    px_py = px[:, None] * py[None, :] # Broadcast to multiply marginals
    # Now we can do the calculation using the pxy, px_py 2D arrays
    nzs = pxy > 0 # Only non-zero pxy values contribute to the sum
    return np.sum(pxy[nzs] * np.log(pxy[nzs] / px_py[nzs]))
```

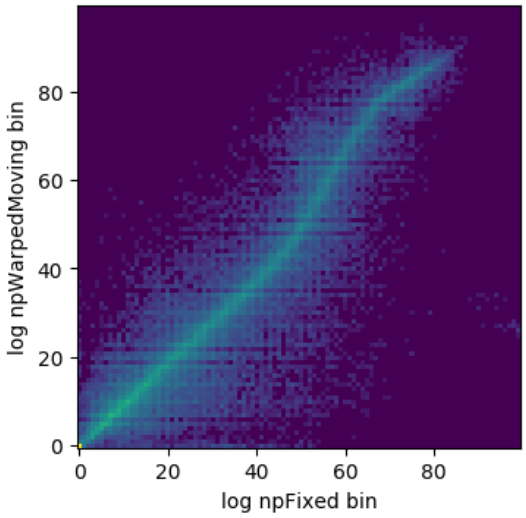
The joint (log) 2D histograms in our baseline and LDOF results are shown in Figure 5-11, and Figure 5-12 shows the evaluation of the baseline models and LDOF according to the four metrics.



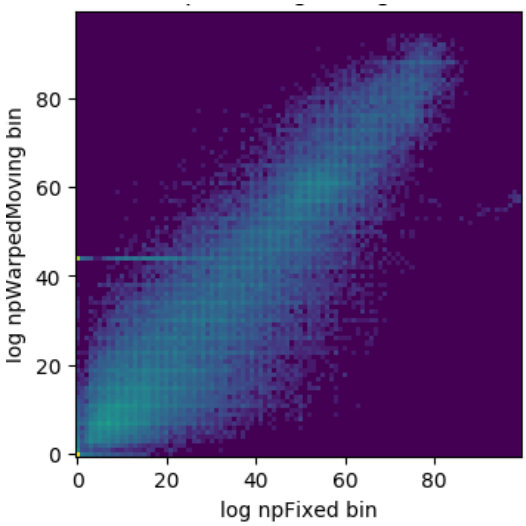
(a) Multiscale Horn-Schunk



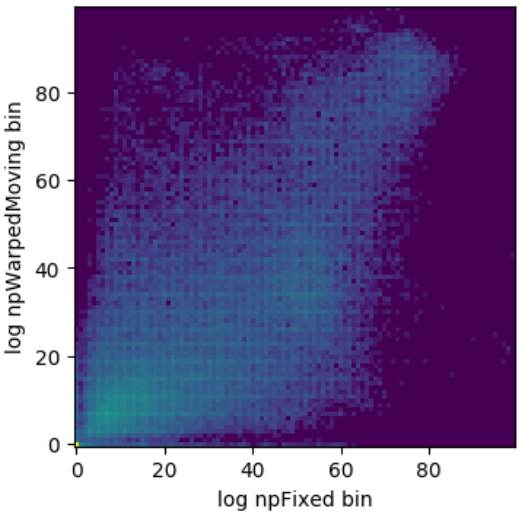
(b) Demon



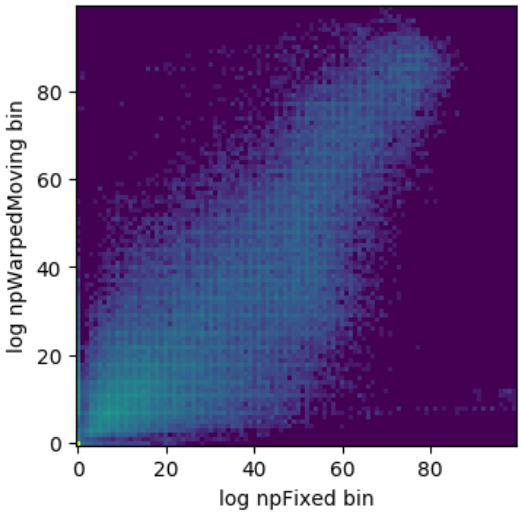
(c) Symmetric Demon



(d) BSpline + LBFGS



(e) Bspline + RSGD



(f) LDOF

Figure 5-11: Joint 2D (log) histogram of I_1 and warped I_2

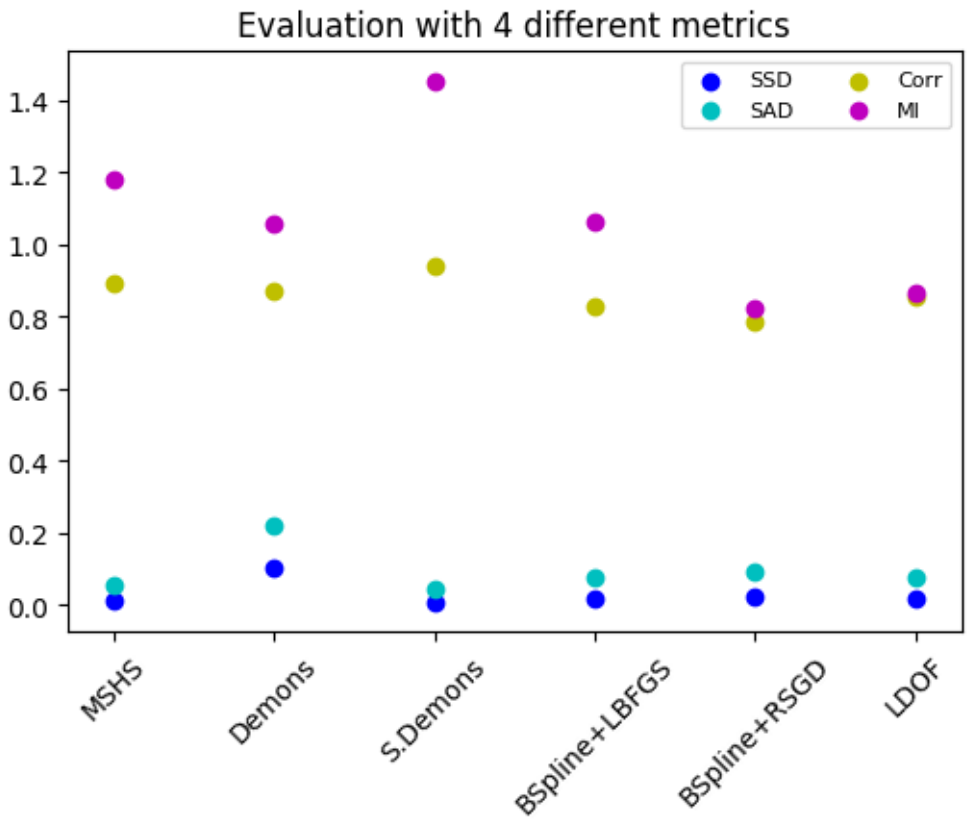


Figure 5-12: Evaluation using 4 metrics: SSD, SAD, Normalized cross-correlation, Mutual Information. The scores are inconsistent due to its dependency on the choice of metric used in objective functions.

This comparison using different metrics demonstrate their contrasting measure of registration quality. One way to overcome such inconsistency is to manual label correspondences between two images, and use them as the ground truth. However, this method is time-consuming and requires a medical specialist's annotation. More automated way to validate and evaluate registrations is needed to compare the real performances of the algorithms more accurately.

Bibliography

[1] ITK deformable registration v4.

[2] Matlab: Image registration.

[3] Matlab: non-rigid registration tool.

[4] Mutual information as an image matching metric âĂŤ Tutorials on imaging, computing and mathematics.

[5] Ulas Bagci. Medical image registration i (introduction), Spring 2016.

[6] S. Baker, S. Roth, D. Scharstein, M. J. Black, J. P. Lewis, and R. Szeliski. A Database and Evaluation Methodology for Optical Flow. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8, October 2007.

[7] Thomas Brox, AndrÃĺs Bruhn, Nils Papenberg, and Joachim Weickert. High Accuracy Optical Flow Estimation Based on a Theory for Warping. In *Computer Vision - ECCV 2004*, Lecture Notes in Computer Science, pages 25–36. Springer, Berlin, Heidelberg, May 2004.

[8] Thomas Brox and Jitendra Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE transactions on pattern analysis and machine intelligence*, 33(3):500–513, 2011.

[9] W R Crum, T Hartkens, and D L G Hill. Non-rigid image registration: theory and practice. *The British Journal of Radiology*, 77(suppl_2):S140–S153, December 2004.

[10] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013.

[11] Evan Herbst, Xiaofeng Ren, and Dieter Fox. Rgb-d flow: Dense 3-d motion estimation using color and depth. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 2276–2282. IEEE, 2013.

[12] Berthold K. P Horn and Michael J Brooks. The variational approach to shape from shading. *Computer Vision, Graphics, and Image Processing*, 33(2):174–208, February 1986.

[13] Berthold K. P. Horn and Brian G. Schunck. Determining Optical Flow. *Artifical InâĎąligence*, 17:185–203, 1981.

[14] Hans J. Johnson, M. McCormick, L. Ibáñez, and The Insight Software Consortium. *The ITK Software Guide*. Kitware, Inc., third edition, 2013. *In press*.

[15] Dirk-Jan Kroon. B-spline grid, image and point based registration.

[16] D. C. Liu and J. Nocedal. On the limited memory bfgs method for large scale optimization. *Math. Program.*, 45(3):503–528, December 1989.

[17] Enric Meinhardt-Llopis, Javier SÃąnchez PÃľrez, and Daniel Kondermann. Horn-Schunck Optical Flow with a Multi-Scale Strategy. *Image Processing On Line*, 3:151–172, July 2013.

[18] Xavier Pennec, Pascal Cachier, and Nicholas Ayache. Understanding the âĂIJdemonâĂŹs algorithmâĂİ: 3d non-rigid registration by gradient descent. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 597–605. Springer, 1999.

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

[20] Richard Szeliski and James Coughlan. Spline-based image registration. *International Journal of Computer Vision*, 22(3):199–218, 1997.

[21] Jean-Philippe Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998.

[22] I. M. J. Van der Bom, Stefan Klein, Marius Staring, R. Homan, Lambertus W. Bartels, and Josien PW Pluim. Evaluation of optimization methods for intensity-based 2d-3d registration in x-ray guided interventions. In *Medical Imaging 2011: Image Processing*, volume 7962, page 796223. International Society for Optics and Photonics, 2011.

[23] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Symmetric log-domain diffeomorphic registration: A demons-based approach. *Imag. Comput*, pages 754–761, 2008.

[24] Andreas Wedel, Thomas Pock, Christopher Zach, Horst Bischof, and Daniel Cremers. An Improved Algorithm for TV-L 1 Optical Flow. In Daniel Cremers, Bodo Rosenhahn, Alan L. Yuille, and Frank R. Schmidt, editors, *Statistical and Geometrical Approaches to Visual Motion Analysis*, pages 23–45, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.