Fall 2009

Non-rigid image registration evaluation using common evaluation databases

Ying Wei
University of Iowa

NON-RIGID IMAGE REGISTRATION EVALUATION USING COMMON EVALUATION DATABASES

by

Ying Wei

A thesis submitted in partial fulfillment of the
requirements for the Master of Science
degree in Electrical and Computer Engineering
in the Graduate College of
The University of Iowa

December 2009

Thesis Supervisor: Associate Professor Gary E. Christensen

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

MASTER'S THESIS

This is to certify that the Master's thesis of

Ying Wei

has been approved by the Examining Committee for the thesis requirement for the Master of Science degree in Electrical and Computer Engineering at the December 2009 graduation.

Thesis Committee: _____
Gary E. Christensen, Thesis Supervisor

_____
Jon G. Kuhl

_____
Erwei Bai

ACKNOWLEDGEMENTS

I would like to thank Prof. Gary E. Christensen for his patient support and supervision of this work and thesis. I would also like to thank Prof. Jon G. Kuhl and Prof. Erwei Bai in ECE Department for their great support in this project. I sincerely thank my workmate Joo Hyun Song, Jeff Hawley and Kate Rasing for their hard work on this project. Great thanks to Kunlin Cao, Cheng Zhang and Xiujuan Geng for their help and friendship. Special thanks goes to Prof. David Rudrauf, Joel Bruss and Prof. Thomas J. Grobowski at Neurology Department for providing the data used for this work. I would also like to thank Prof. Milan Sonka and the other faculty, staff and students in the Electrical and Computer Engineering department for many informative seminars and lectures.

This work is supported in part by NIH grant EB004126.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1
INTRODUCTION

1.1 Motivation and Background

Image registration is important for many medicine applications, such as image segmentation, multi-modality fusion, functional brain mapping, deformable atlas registration and image guided surgery. Non-rigid image registration is a more general approach than the widely used rigid methods, but requires more complex methodology and computational effort to implement [9]. Various non-rigid registration algorithms are developed by research community today, yet the performance is difficult to assess. Evaluation of the performance is important as it helps to clarify the potential clinical application that a algorithm might serve [23]. However, evaluating the performance of non-rigid image registration algorithms is a hard task since point-wise correspondence from one image to another is not unique, i.e., there is rarely if ever a "Gold Standard" or ground truth correspondence map to judge the performance of a registration algorithm [9]. Comparison of performance of different medical imaging processing algorithms requires the use of standardized or at least rigorous terminology and common methodology for the validation process. And creating test data sets and evaluation statistics enable generalizable conclusions to be drawn [23]. As we know, no registration algorithm will perform the same for all types of input images [21], for example, one algorithm may perform very well for MR registration but poorly for CT. And in general no algorithm will perform best on all evaluation statistics. Thus, different

evaluation databases and evaluation statistics (methods) are needed. An increasing number of researchers are developing evaluation databases and evaluation statistics for evaluating image registration algorithms. The website (http://www.vmip.org/) set up by Pierre Jannine directs people to papers and references about validation and evaluation in medical imaging processing, and a list of validation data sets. The subjects discussed on this website are not only for image registration, but for all medical image processing methods. When it comes to evaluation of medical image registration, many evaluation statistics have been developed and widely used. As the number of all statistics are too large to quote them all, we refer readers to [32], [40], and [17] for an overview on them. These papers list and summarize many currently available image registration performance evaluation methods. Generally speaking, evaluation statistics can be divided to two goups. One kind is the similarity-based statistics which measures how similar two images are. There are some examples: (1) Overlap. This kind of statistics all measure how well the labeled volumes or surfaces of source image and target image are agree or disagree to each other before and after registration. For more information about overlap measures, see the paper by Crum et al. [16]. Some overlap statistics are listed here: relative overlap (included in this work), DICE coefficient, target overlap ([26]); (2) Landmark error. These statistics measure the distance between deformed landmarks and corresponding target landmarks. Distance between two point sets can be measured using Euclidean distance, Hausdorff distance, closest distance or any other suitable metric. Readers can see distance error in [26] as an example; (3) Intensity based error. These measure intensity

difference between deformed and target intensity images. Examples of these errors include intensity variance, mutual information and average volume method ([22]). And the other kind of evaluation statistics is the transformation based statistic related to transformation properties. The transformation (deformation field) error, for example. is modeled to measure different properties of the deformation field. Examples of these statistics include inverse consistency error, transitivity error and the Jacobian of the transformation ([11]) which characterizes the possible singularities of the deformation field especially zero-crossing([22]).

Further, there have been a few attempts made to objectively evaluate and compare the performance of image registration algorithms using standard evaluation databases and statistics [6]. Several significant prior studies should be mentioned here. A recent work is conducted by Yassa and Stark [39], who analyzed how accurate cross-participant alignment was used for evaluation within the medial temporal lobe (MTL). Twenty structural MRI data sets were collected in this study. They evaluated the performance of three fully non-rigid registration algorithms and two evaluation statistics were used to assess the performance of each algorithm. Klein et al. [26] evaluated non-rigid registration algorithms performance using human brain MRI. This study is the largest evaluation of nonlinear deformation algorithms applied to brain image registration conducted to date. Fourteen algorithms from laboratories around the world were evaluated using 8 different error measures. Other research groups provide the community with images to register and then evaluate the results. The "Retrospective Image Registration and Evaluation Project" [37] led by Jay West

Fitzpatrick of Vanderbilt University took this approach to evaluate inter-modality registration algorithms. A common set of images were used to evaluate the performance of registration algorithms. Researchers registered the images with their own registration algorithms and then send an ASCII file containing the original and transformed points back to Vanderbilt. Registration algorithms were evaluated using the target registration error. Another non-rigid registration evaluation project entitled "Retrospective Evaluation of Inter-subject Brain Registration" [22] led by P. Hellier of IRISA/INRIA-CNRS Rennes, France evaluate 6 registration algorithms using 18 brain image volumes. The evaluation statistics (measures) used were divided to global measures and local measures. People around the world participated in this projects by registering the images with their own registration algorithms and sending the resulting transformations back to the home site for analysis. They chose an arbitrary subject as the reference subject and compared results of registration by registering other images to this target image. Castillo et al. [8] evaluated deformable image Registration (DIR) spatial accuracy using large sets of expert-determined landmark point pairs. Each of their data sets has associated with it a coordinate list of anatomical landmark point sets which serve as a reference of evaluating DIR spatial accuracy within the lung. They provide published DIR spatial accuracy results on their website (http://www.dir-lab.com). Results are reported as mean 3D Euclidean magnitude distance between calculated and reference landmarks. Castillo et al. mentioned the need for actual patient image data for evaluation since other kinds of reference standards like synthetic images [28] and phantoms [36] lack sufficient

realism to provide credible validation of registration spatial accuracy for use in the clinical setting. Besides these image registration evaluation studies, another important project is the VALMET software tool for assessing and improving segmentation developed by Guido Gerig et al. [20] (www.ia.unc.edu/public/valmet/). The VALMET software is a software tool for measuring and visualizing the differences between multiple corresponding medical image segmentations over four statistics.

The Non-rigid Image Registration Evaluation Project (NIREP) [9] was started to develop a standardized set of common databases, evaluation statistics and software tool for performance evaluation of non-rigid image registration algorithms. NIREP provides new statistics for evaluating registration performance as well as well-known statistics involved in previous projects. Under our model, NIREP software users can evaluate the performance of different non-rigid registration algorithms on NIREP database downloaded from NIREP website. This model has the advantage of standardizing the evaluation database and distributing the evaluation results. It provides researchers with a tool to compare the performance of multiple registration algorithms so they can make an informed decision regarding the best algorithm for their specific application. The results and analysis will be disseminated through publications and a central web site (http://www.nirep.org). The goal of the work in this paper is to build up common image databases for rigorous testing of non-rigid image registration algorithms, and compare their performance by a diverse set of evaluation statistics on our multiple well documented image databases. The well documented databases as well as new evaluation statistics have been and will be released to public research

community. The tests in this paper were used to evaluate the performance of image registration algorithms following NIREP framework with respect to their transformation properties, agreement with human experts, and other indirect performance tests. This paper provides a complete and accurate reporting of evaluation tests so that others are able to get access to these results and make a comparison of registration algorithms they concerned in their specific use. Moreover, this work followed the recommendations of the Standards for Reporting of Diagnostic Accuracy (STARD) initiative to disclose all relevant information for each non-rigid registration validation test.

The work in this thesis is built upon these earlier projects mentioned above and has the following important features: (1) our experiments were conducted using multiple well characterized databases. This helps to show whether the performance of different registration algorithms are affected by the choice of subject populations with respect to different statistics; (2) Instead of choosing an arbitrary subject as the reference subject, our study used all the images as the reference subject. This helped avoid the influence of the choice of the reference subject [37]; (3) This thesis provides a detailed description of the NIREP evaluation framework, which is generalizable to new subject populations, new registration algorithms and new statistics; (4) We investigated using fMRI data to evaluate how well registration algorithms registered anatomical MR images. We also discussed the impact of non-rigid anatomical registration on the variability of the functional data. The "Retrospective Evaluation of Inter-subject Brain Registration" [22] also mentioned that they plan to focus on

functional data to see what the impact of non-rigid anatomical registration has on the variability of the functional data; (5) We are building an exportable software tool that will have the non-rigid image registration evaluation statistics built into it. This is in contrast to these previous projects that collected registration transformations from external participants and performed the analysis internally. (6) The model of STARD documentation is followed to describe all relevant information for each non-rigid registration evaluation test. This meets the need to archive registration algorithm performance so that people can choose the best registration algorithm based on their own needs in the future. Archiving registration results can help drive future innovation in non-rigid image registration.

1.2 Outline

This thesis consists the following parts. Chapter 2 describes the NIREP framework, the two evaluation databases developed for this project, five non-rigid image registration algorithms, the statistics used to evaluate registration algorithms, and the STARD documentation.

Chapter 3 describes evaluation results, discusses the evaluation result by fMRI, and presents the STARD documentation for the evaluation result.

Chapter 4 discusses and analyzed the evaluation results obtained from this work. Chapter 5 summarizes this thesis work. Finally, chapter 6 states some problems encountered and explores future extensions.

CHAPTER 2
METHODS

2.1 NIREP framework

Figure 2.1 shows the NIREP framework for evaluating registration performance. The evaluation framework consists of the following steps: (0) Obtain a non-rigid image registration algorithm to evaluate. (1) Download the evaluation database, the evaluation software, STARD documentation describing the evaluation database, and pre-computed results from NIREP website. (2) Register images from the NIREP evaluation database using the registration algorithm to evaluate and store the computed transformations for analysis. (3) (Optional) Apply transformations to evaluation database to generate deformed images (MRI, CT, fMRI, DTI etc.), segmentations, point sets (e.g., landmarks) etc. (4) Use NIREP software to evaluate registration performance using multiple evaluation statistics. (5) Compare registration performance to that of other algorithms using data downloaded from web site. (6) Document results of analysis by completing STARD documentation. (7) (Optional) Upload results and STARD documentation to NIREP website to share with other investigators.

The thesis presents an example registration evaluation by common databases using this framework to evaluate non-rigid image registration algorithms. Two NIREP neuroanatomical evaluation databases were established in this work. 22 MR data sets from these two databases, 6 registration algorithms (1 rigid and 5 non-rigid), 6

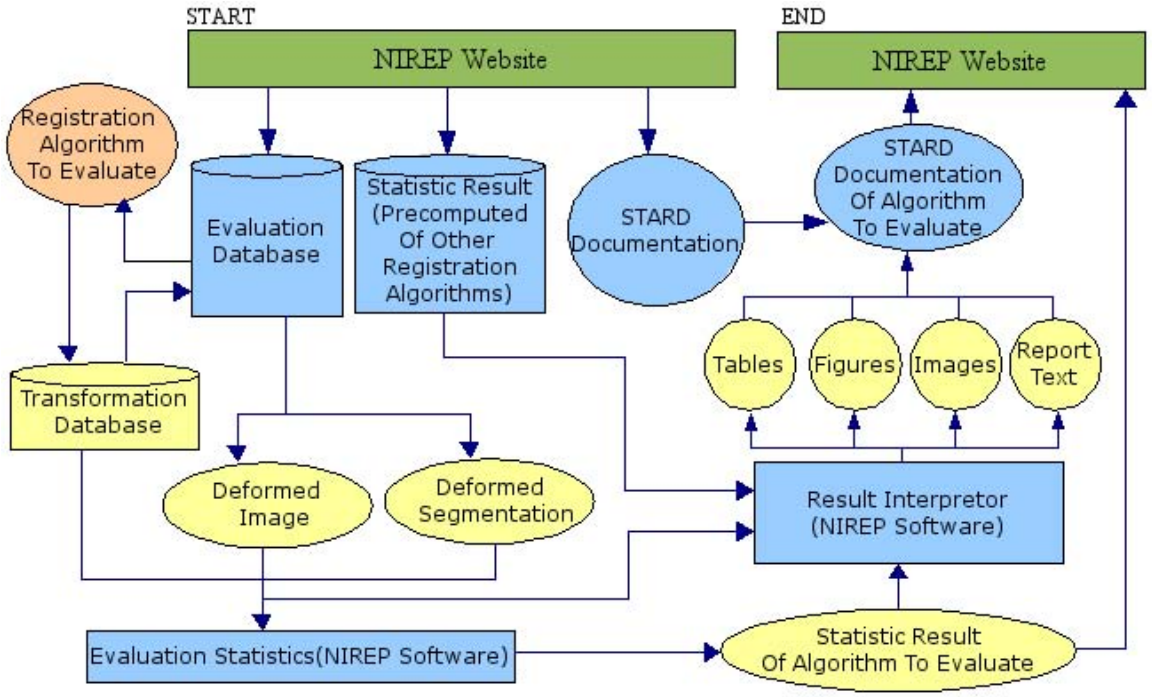


Figure 2.1: NIREP framework.

evaluation statistics were used as a baseline result of NIREP.

2.2 Database Acquisition and Distribution

2.2.1 NIREP NA0

Databbase Acquisition

The initial evaluation database consists of a population of 16 richly annotated 3D MR image volumes corresponding to 8 normal adult males and 8 females. These data sets were selected from a database of healthy right-handed individuals housed in the Human Neuroanatomy and Neuroimaging Laboratory. The demographics of these data sets are shown in Table 2.1. The males have a mean age of 32.5 years, standard deviation of 8.4 years and range in age from 25 to 48. The females have a

Table 2.1: Clinical demographic characteristics of the study population of NA0.

Label	Age	Gender	Race	Ethnic Category	Handedness
na01	43	Male	White	Non Hispanic	+95
na02	48	Male	White	Non Hispanic	+95
na03	28	Male	White	Non Hispanic	+85
na04	28	Male	Asian	Non Hispanic	+100
na05	32	Male	Unknown	Hispanic	+100
na06	27	Male	White	Non Hispanic	+80
na07	29	Male	White	Non Hispanic	+65
na08	25	Male	White	Non Hispanic	+100
na09	26	Female	White	Non Hispanic	+100
na10	27	Female	Asian	Non Hispanic	+100
na11	36	Female	White	Non Hispanic	+95
na12	26	Female	White	Non Hispanic	+85
na13	24	Female	Unknown	Hispanic	+100
na14	28	Female	White	Non Hispanic	+80
na15	30	Female	Black	Non Hispanic	+100
na16	41	Female	White	Non Hispanic	+100

mean age of 29.8 years, standard deviation of 5.8 and range in age from 24 to 41.

Figure 2.2 shows an example of T1 image and the segmentations associated with the MR data sets.

The 16 MR data sets of the NA0 database were segmented into 32 Region of Interests (ROIs). Fig 2.3 shows an example of the segmentations of these MR data sets generated by this project. Table 2.2 associates a label with each ROI. Note that odd and even numbered objects correspond to ROIs on the left and right side of the brain, respectively.

Some of these data sets and their segmentations appear in a new atlas by

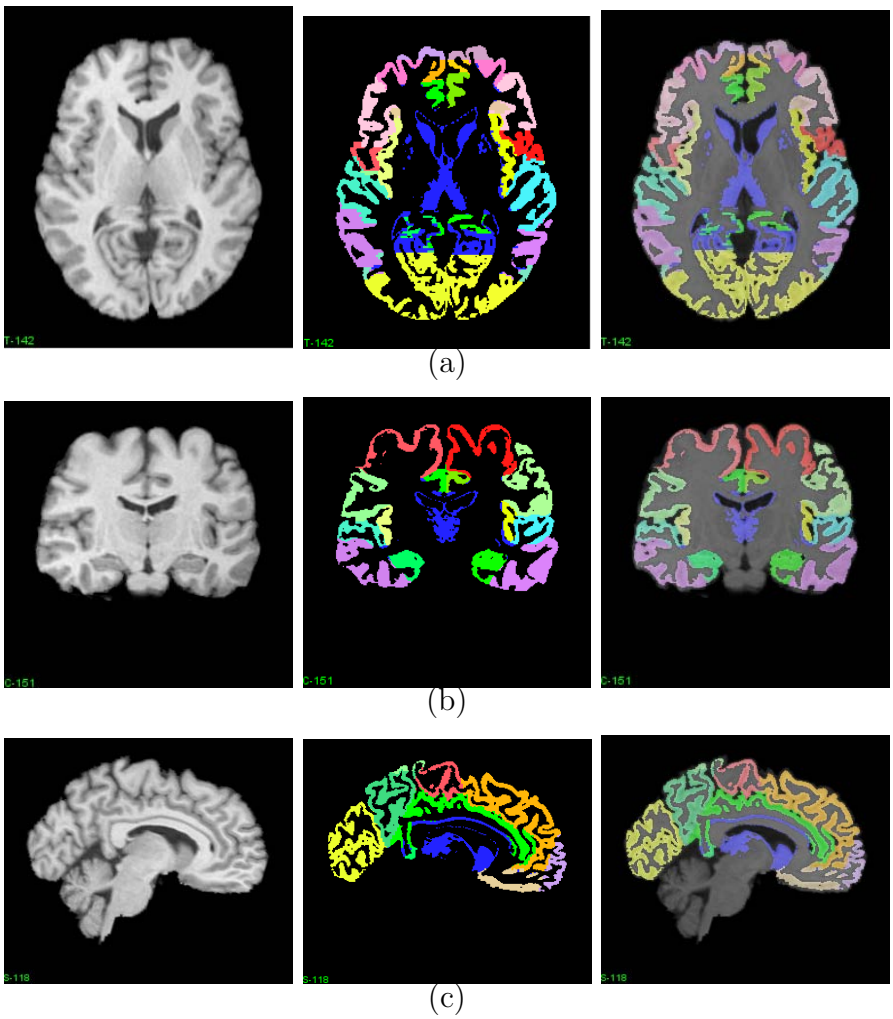


Figure 2.2: The na01 data set from NIREP NA0 database. Shown are the T1 image (left column), the segmentation image associated with the na01 data set (Middle column), and the segmentation overlapped on the T1 image (right column) viewed in (a) transverse, (b) coronal and (c) sagital slices.

Table 2.2: Regions of Interest (ROI) in the NA0 evaluation database. The average volume for each ROI is reported in units of voxels.

	ROI	ave volume $\times 10^5$		ROI	ave volume $\times 10^5$
1	L occipital lobe	0.87	2	R occipital lobe	0.93
3	L cingulate gyrus	0.42	4	R cingulate gyrus	0.45
5	L insula gyrus	0.22	6	R insula gyrus	0.21
7	L temporal pole	0.28	8	R temporal pole	0.31
9	L superior temporal gyrus	0.45	10	R superior temporal gyrus	0.39
11	L infero temporal region	1.0	12	R infero temporal region	1.0
13	L parahippocampal gyrus	0.35	14	R parahippocampal gyrus	0.34
15	L frontal pole	0.17	16	R frontal pole	0.18
17	L superior frontal gyrus	0.79	18	R superior frontal gyrus	0.78
19	L middle frontal gyrus	0.67	20	R middle frontal gyrus	0.64
21	L inferior gyrus	0.30	22	R inferior gyrus	0.32
23	L orbital frontal gyrus	0.46	24	R orbital frontal gyrus	0.4
25	L precentral gyrus	0.62	26	R precentral gyrus	0.62
27	L superior parietal lobule	0.64	28	R superior parietal lobule	0.60
29	L inferior parietal lobule	0.78	30	R inferior parietal lobule	0.82
31	L postcentral gyrus	0.45	32	R postcentral gyrus	0.42

Dr. Hanna Damasio that illustrates the normal neuroanatomy of the human brain, and have been used in several publications about the morphometric analysis of the normal human brain [18]. Dr. Thomas J. Grabowski re-examine the segmentations and approve the anatomical definitions. The criteria for inclusion and exclusion are documented using the STARD documentation.

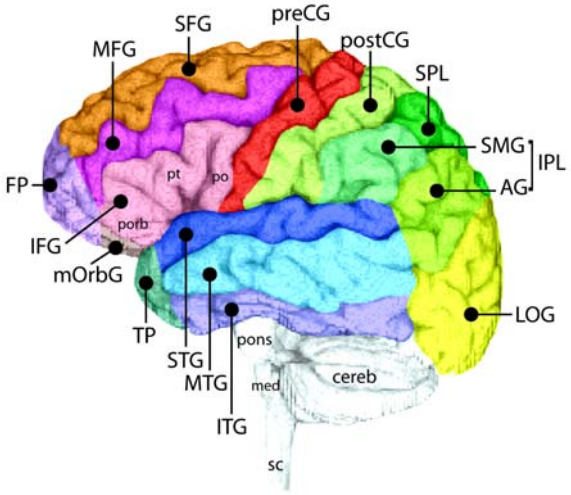


Figure 2.3: The segmentations avaliable in NIREP NA0 (Detail of figure 2 of chapter 2, H. Damasio, "Human Brain Anatomy in Computerized Images," 2nd ed., 2004, Oxford University Press, In press). Shown are typical segmentations available to this project and include: Cerebrum: The cerebellum, hypothalamus, and brain stem are not segmented; Left and right hemispheres: Frontal Lobe: Frontal Pole, Superior Frontal Gyrus, Middle Frontal Gyrus, Inferior Frontal Gyrus, Orbital Frontal Gyrus, Precentral Gyrus; Parietal Lobe: Postcentral Gyrus, Superior Parietal Lobule, Inferior Parietal Lobule; Temporal Lobe: Temporal Pole, Superior Temporal Gyrus (including Heschl's Gyrus (Primary Auditory Cortex) and Planum Temporale), Infero-Temporal Region, Parahippocampal Gyrus (including the Amygdala and Hippocampus); Occipital Lobe; Cingulate Gyrus; Insula.

Tissue segmentation was performed by Joel Bruss BA, under the supervision

of neurologist Dr. Grabowski.Cortical parcellation was performed by Joel Bruss BA and John Allen PhD, under the supervision of neurologists Hanna Damasio MD and Thomas J. Grabowski, MD. Combination of tissue segmentation and cortical parcellation was performed by doctoral student Xiujuan Geng, and research asistant Joel Bruss BA, under the supervision of Thomas J. Grabowski, MD.

2.2.2 NIREP NA1 Database Acquisition

2.2.2.1 Database Description

NIREP NA1 database consists of a population of 18 MRI data sets consisting of 9 normal adult males and 9 females. These data sets were acquired in the Human Neuroanatomy and Neuroimaging (HNN) Laboratory, The University of Iowa. The demographics of these data sets are shown in Table 2.3. All the subjects here are right handed and native English speakers. This database consists of structural and functional MRI and 57 segmented regions of interests.

The study population was a consecutive series of participants recruited from: (1) The University Hospital setting, recruited through local advertising; (2) Local communities in Iowa, recruited through newspaper advertising (and with special targeting of minority populations); (3) A dataset of approximately 80 normal control subjects who agree to be contacted for future studies under IRB# 2000303002 - Mechanisms of Perimetric Variability - Michael Wall, PI. and (4) College students at the University of Iowa, through local advertising; (5) Referral from colleague - Patients of Michael Wall, MD. Dept. of Neurology, who is a co-investigator on this project; (6) Other-Word of mouth. The inclusion/exclusion criteria for MRI included in this

Table 2.3: Clinical demographic characteristics of the study population of NA1.

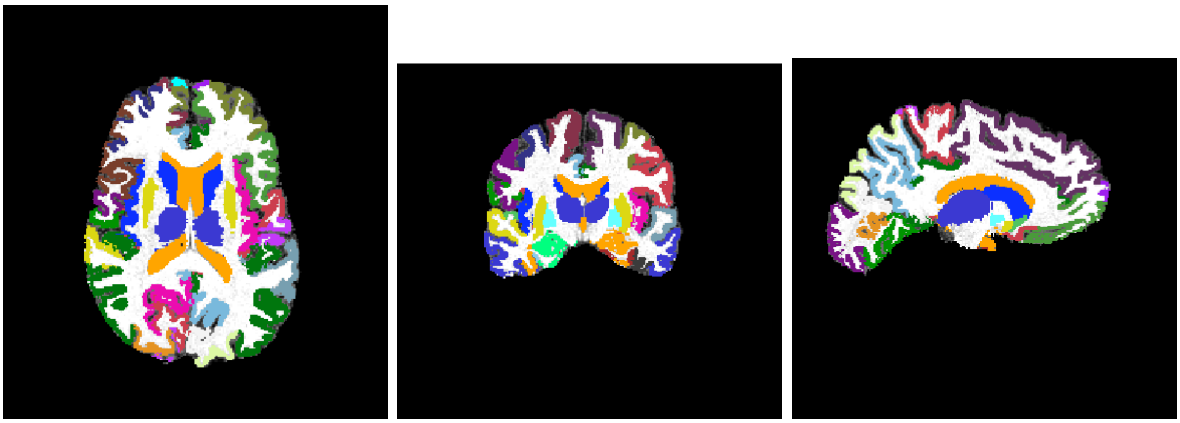
Label	Age	Gender	Race	Handedness
3065	35	Male	White	+100
3362	23	Male	White	N/A
3368	26	Male	White	N/A
3402	23	Male	Hispanic	+100
3403	32	Male	African Am.	+95
3407	28	Female	White	+100
3413	26	Female	White	+100
3414	31	Male	African Am.	+90
3424	24	Female	White	+100
3425	23	Male	White	+95
3463	24	Male	White	+90
3468	36	Female	White	+100
3474	30	Female	White	+100
3479	29	Female	White	+80
3481	39	Male	White	+95
3489	56	Female	Asian/Pacific	+100
3491	27	Female	Islander	+100
3493	22	Female	African Am.	+95

database are: (1) $Age > 18 years$; (2) No history of developmental, neurological or psychiatric disease; (3) Normal visual acuity (can read text in a book at arms length without glasses); (4) No history of medical disease or medication which might affect cerebral blood flow; (5) No factor which contraindicates MR scanning, including pacemaker, pacemaker wires, implanted cardiac defibrillator, Neurostimulator, aneurysm clip, or any electronic implant, inner ear surgery, weight over 136 kg (300lb), metal embedded in soft tissue or in the eye, prosthetic eye, or claustrophobics; (6) Not pregnant; (7) Not non-English speaker; (8) Do not smoke tobacco; (9) No non-removable body piercing; (10) No neurological conditions of any of the following: stroke, severe head trauma (motor vehicle accident, loss of consciousness, alteration of consciousness or memory loss), tumor, meningitis, encephalitis, seizure disorder, severe migraine, dementia, epilepsy, any other neurological condition which may contribute to cognitive impairment; (11) No developmental disabilities, including dyslexia and a learning disability; (12) No medical conditions of the following: severe hypertension, severe thyroid dysfunction, severe anemia and/or sickle cell disease, renal failure, heart disease, diabetes; (13) Not use anti-depressants; (14) No history of depression; (15) No regular use of antihistamine; (16) Not an employee of the Neurology Department or a medical students rotating in Neurology; (17) No conditions of any of the following: employee of the PI or employee of a research team member, individual supervised by PI or supervised by member of research team, individual subordinate to the PI or subordinate to any member of the research team, student or trainee under the direction of the PI or under the direction of a member of the research team; (18) Not

incompetent or have limited decision-making capacity on initial enrollment into the study; (19) No change of capacity to consent over the course of the study; (20) Not prisoner.

2.2.2.2 Data Segmentation

In the NA1 database, 18 MR datasets were segmented into 57 gray matter regions of interest(ROI). Figure 2.4 shows an example of the segmentations associated with the MR datasets.



(a) (b) (c)

Figure 2.4: The 3065 data set from NIREP NA1 database. Shown is the segmentation image (color) overlapped on the T1 image (gray intensity) viewed in (a) transverse, (b) coronal and (c) sagital slices.

Table 2.4 associates a label with each ROI in NA1 database.

Gray matter parcels were taken from the output of FreeSurfer (FS) (Based on what it interprets to be GM). Specifically, the Desikan-Killiany atlas was used [18]. The final parcel mask was produced by hand editing the result parcel produced by

Table 2.4: Regions of Interest (ROI) in the NA1 evaluation database. "ROI number" refers to the value of each ROI in the segmentation. "name" (e.g. Lcun) refers to the name of each ROI.

ROI number	ROI name	ROI number	ROI name
0	Background	1	Still unsigned regions
2	Left Cuneus	3	Right Cuneus
4	Left Lateral Occipital Gyrus	5	Right Lateral Occipital Gyrus
6	Left Lingual Gyrus	7	Right Lingual Gyrus
8	Left Calcarine (pericalcarine) Region	9	Right Calcarine (pericalcarine) Region
10	Left Cingulate Gyrus	11	Right Cingulate Gyrus
12	Left Insular Cortex	13	Right Insular Cortex
14	Left Temporal Pole	15	Right Temporal Pole
16	Left Superior Temporal Gyrus	17	Right Superior Temporal Gyrus
18	Left Heschl's Gyrus	19	Right Heschl's Gyrus
20	Left Middle Temporal Gyrus	21	Right Middle Temporal Gyrus
22	Left Inferior Temporal Gyrus	23	Right Inferior Temporal Gyrus
24	Left Fusiform Gyrus	25	Right Fusiform Gyrus
26	Left Entorhinal Cortex	27	Right Entorhinal Cortex
28	Left Parahippocampal Gyrus	29	Right Parahippocampal Gyrus
30	Left Frontal Pole	31	Right Frontal Pole
32	Left Superior Frontal Gyrus	33	Right Superior Frontal Gyrus
34	Left Middle Frontal Gyrus	35	Right Middle Frontal Gyrus
36	Left Inferior Frontal Gyrus	37	Right Inferior Frontal Gyrus
38	Left Orbital Gyri	39	Right Orbital Gyri
40	Left Paracentral Lobule	41	Right Paracentral Lobule
42	Left Precentral Gyrus	43	Right Precentral Gyrus
44	Left Postcentral Gyrus	45	Right Postcentral Gyrus
46	Left Superior Parietal Lobule	47	Right Superior Parietal Lobule
48	Left Inferior Parietal Lobule	49	Right Inferior Parietal Lobule
50	Left Precuneus	51	Right Precuneus
52	Caudate.	53	Putamen.
54	Globus Pallidus.	55	Nucleus Accumbens
56	Thalamus	57	Lateral Ventricles

FS. For Thalamus, the FS output was edited by hand to fix mistakes. The Insulas were traced by hand and then subtracted from the FS parcel ctx-"hemi"-unknown. After the Insula, Thalamus, and Ventricle were traced, the sum total was subtracted from the ctx-"hemi"-unknown of the FS output parcel. The remainder of this parcel was assigned ROI name as "ctx-lh-unknown, ctx-rh-unknown".

(a) (b) (c)

Figure 2.5: The graymatter image and the parcel image used to generate segmentations for NIREP NA1. (a) The graymatter image is overlayed by graymatter mask (threshold=0.7). Yellow color refers to the graymatter mask. We can see that the floating points in the graymatter image which have values less than 0.7 are not included in the graymatter mask. (b) Parcels which combine the results from FreeSurfer and work by hand (c) parcels overlapped by graymatter mask. This shows some parcel points(with color) but not graymatter(white) need to be deleted.

After the final parcel mask was produced, next task was to dilate the labeled segmentation produced by FS to match the "true" GM (> 0.7) to get the final segmentations. It was decided to leave ROI labeled 52-57 (basal gray regions and ventricles) alone, but to expand all other ROI (including "Unassigned Gray") and to add these regions back later. That means, 51 ROI were considered here, 25 in left and 25 in right, plus unassigned gray. To expand a parcel to the graymatter

mask, points in the parcel image which were not in graymatter mask (detected by overlapping graymatter image on parcel image) were first deleted. Second, each ROI was extracted and graymatter points which had not been assigned any ROI value yet were extracted too. Then the distance from each selected graymatter point to each ROI was calculated and the nearest ROI's value was assigned to the point. The threshold of the distance was 6. If the distance from a graymatter point to any ROI was larger than 6, no value was assigned to it. Figure 2.6 shows the procedure described above to generate the segmentations for NIREP NA1 database.

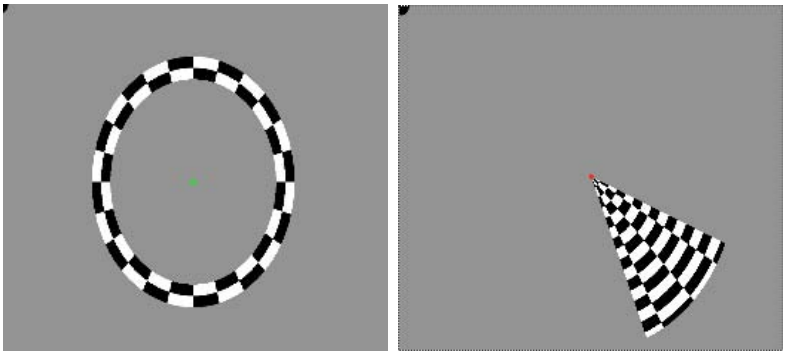
(a) (b) (c)

Figure 2.6: The procedure to generate the segmentations for NIREP NA1 (a) shows a general view of the overlap of colored parcel image(after delete some redundant points) and graymatter image(white). Parcel image is above graymatter image.We can see here are lots of white points, which are graymatter without any ROI value yet; (b)shows the overlap of colored parcel image and graymatter image after processing. Here we can see most points have been assigned a ROI value and have a color now; As 52-57 ROI have not been considered this time, near the center of the brain still have some unsigned graymatter; (c) shows the overlap of three images from top to bottom: colored parcel image(before processing), light colored parcel image(after processing), and graymatter image; This can easily see those points assigned a ROI's value by this program processing.

2.2.2.3 Functional MRI Acquisition

Blood-oxygen-level dependent (BOLD) is the MRI contrast of blood deoxyhemoglobin first discovered in 1990 by Dr. Seiji Ogawa [30] who also recognized the potential importance of BOLD for functional brain imaging with MRI. Almost all fMRI research use BOLD as the method for determining where activity occurs in the brain as the result of various experiences. Since BOLD-weighted fMRI can be used to show regions of brain activity, it can also be used to map regions of the visual cortex, that is, retinotopy. Retinotopy allows stimulation of areas in the visual field to correspond to specific areas of the striate cortex, which is detectable by BOLD-fMRI. Therefore stimulation across the visual field will create a detectable pattern of stimulation across the primary visual cortex.

In NA1 database, T1-weighted structural MRI data were accompanied by fMRI data for each data set. fMRI data were collected for this project and used to evaluate registration performance by correlating structural-functional correspondence. To collect fMRI data, retinotopic experiments were performed. Stimulation in retinotopic experiment here consisted of both expanding rings and rotating wedges composed of black and white stripes that moved laterally in a flowing movement and reversed direction periodically. A fixation point for the subject was located in the center of patients visual field. The fixation point was a small dot that reversed colors between red and green and the subjects were asked to count the number of times the fixation dot changed color in each run and try to figure out how many times the dot changed color on average. These two experiments are shown in Figure 2.7.



(a) (b)

Figure 2.7: Retino experiment. (a) eccentricity (b) polar angle.

2.2.2.4 Registration Preprocessing

Several preprocessing steps were performed on NA1 database after data sets were collected. The first preprocessing step was reorientation of image data. ITK needs to specify the orientation of the output image if they are stored in Analyze format. Analyze 7.5 only supports three types of orientation: RPI, RIP, and PIR. The orientation used here was RPI- voxels ordered from right to left stored a row, rows ordered from posterior to anterior stored a slice, and slices stored from inferior to superior stored a volume, as shown in Figure 2.8:

The second step was histogram normalization. As in practice, MRI images require intensity normalization to equalize the intensities of the image, while CT images do not [11]. Figure 2.9 shows the histogram of all original NA1 T1 images before histogram normalization. We can see there is a shift among the maximum of white matter of subjects.

A simple but effective method to intensity normalization MRI data is to com-

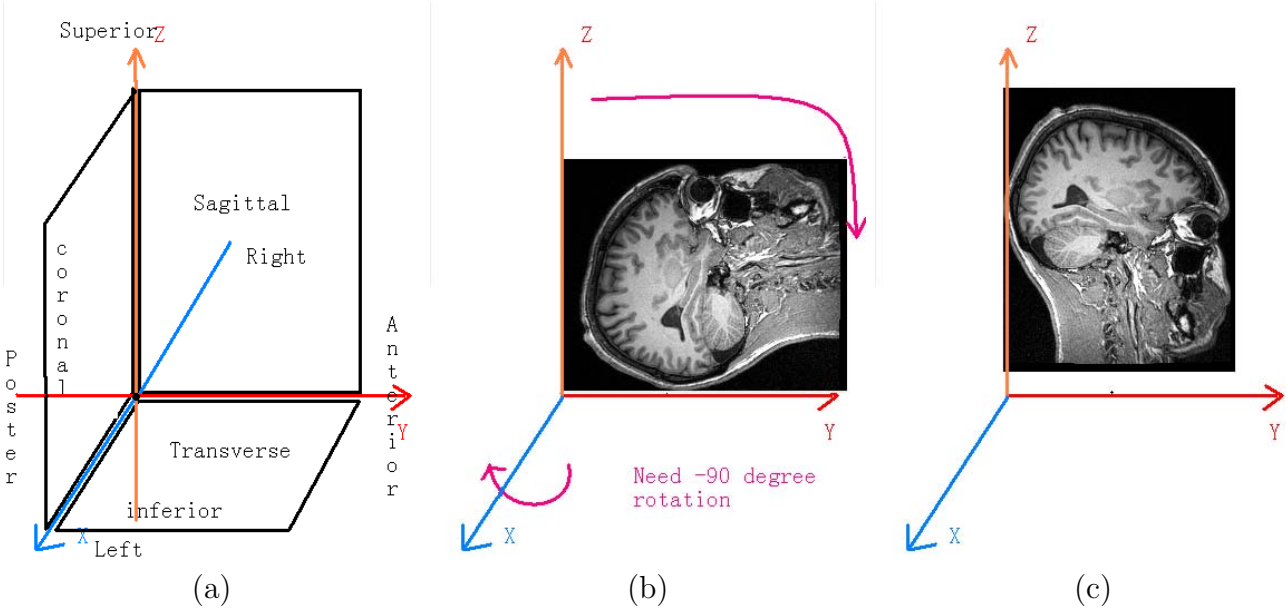


Figure 2.8: The orientation problem with the data sets images(a) coordinates in Analyze 8.1 (b) data sets images before rotation (c)data sets images after rotation

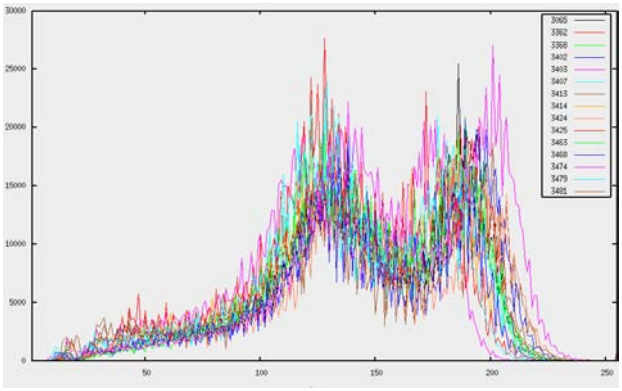


Figure 2.9: Histogram of the original NA1 T1 images before histogram normalization. The x axis is the voxel intensity, and the y axis is the number of voxels. Notice that there is a shift in the tail of the histogram as mentioned before. Notice that the maximum of the white matter of the leftmost histogram and rightmost histogram has a shift of 30 (maximum of WM of leftmost's intensity is 200, and that of rightmost's intensity is 230).

pute the histogram of the two images. Here the axis of one histogram was scaled so that the gray and white matter maximums match those of the other histogram. And then scaled histogram was applied to the image. All data sets in NA1 were normalized using GM mean = 130 and WM mean = 200. Histograms of normalized T1 images are shown in Figure 2.10, and T1 images before and after histogram normalization are shown in Figure 2.11.

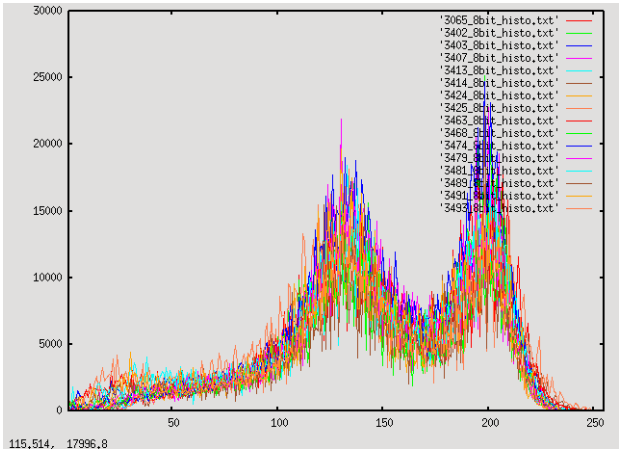
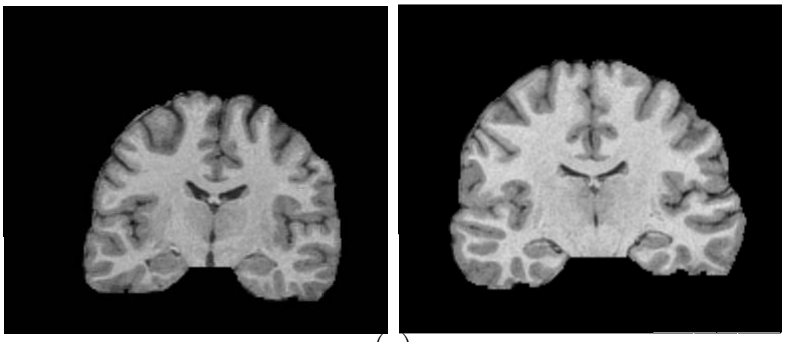
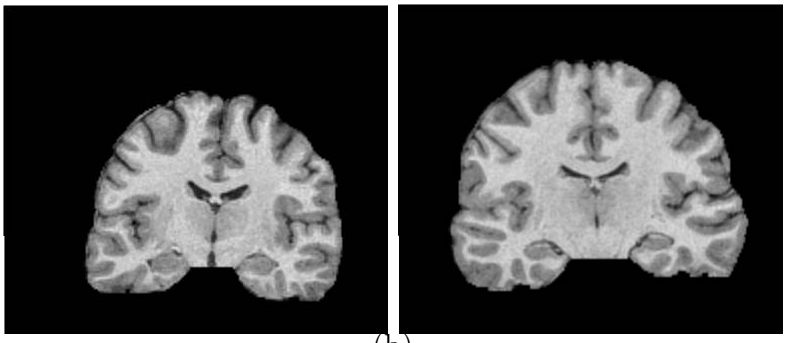


Figure 2.10: Histograms of Normalized T1 images

We also applied histogram normalization on the T2 images. In NIREP NA1 database, T2 Axial (70 slice) volumes, normalized for Gray/White peaks (similar to what was done for the T1 volumes), were saved as 8-bit images. After masking for the cerebrum, a histogram of the T2 data was calculated. Using Gnuplot we chose the peak for the GM and WM (the first two peaks of the histogram) and normalized the output. In most cases, the two peaks were well separated and easy to identify



(a)



(b)

Figure 2.11: Normalized T1 images: 3403 (left) and 3474 (right) from NA1 database.(a) before normalized (b) after normalized.

but, in a few cases, the peaks were pushed together and more difficult to determine. Histograms of normalized T2 images are shown in Figure 2.12, and T2 images before and after histogram normalization are shown in Figure 2.13.

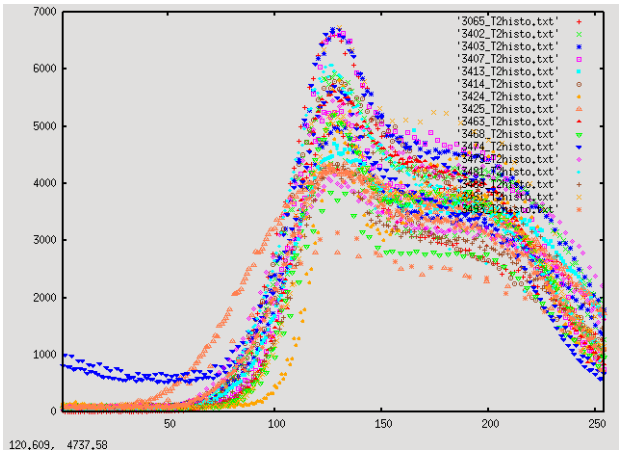
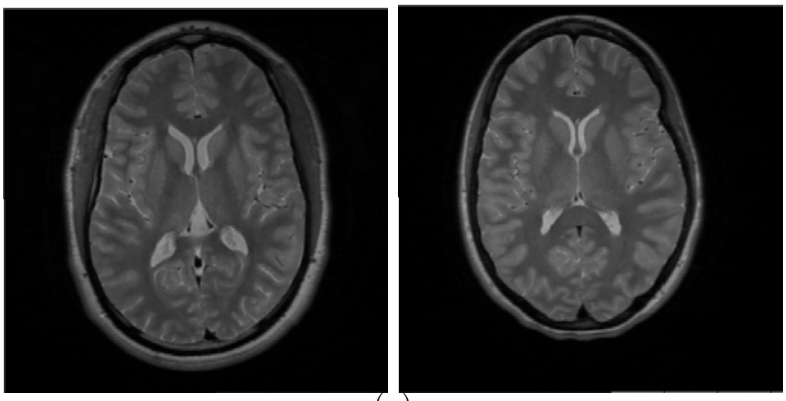


Figure 2.12: Histograms of Normalized T2 images

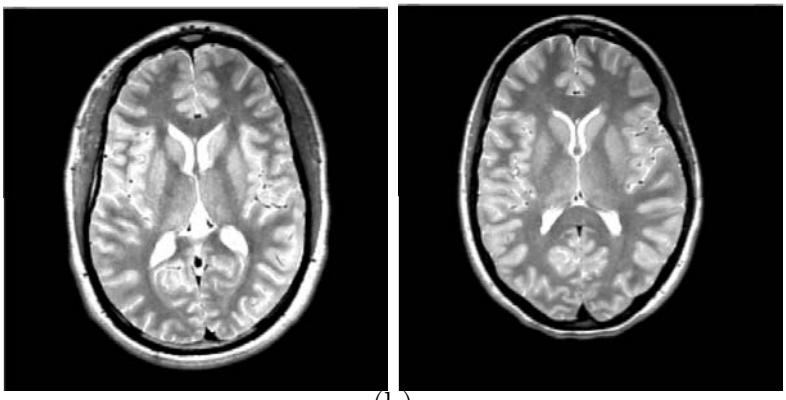
The third step was to determine the AC location. The reference location in the target space of the AC point was :(126,103,116). And the AC point of every data set was moved to this location.

2.2.3 XNAT for NIREP Data Distribution

XNAT central is a database for sharing neuroimaging and related data with select collaborators or the general community through the world wide web [29]. XNAT Central is a service of the Neuroinformatics Research Group at Washington University School of Medicine. It is built using the XNAT data management platform. In XNAT central, in addition to downloading the data sets, people are able to view description



(a)



(b)

Figure 2.13: Normalized T2 images: 3403 (left) and 3474 (right) from NA1 database.(a) before normalized (b) after normalized.

of image database and demographic related to each data set too. Thus XNAT central is a powerful tool for database distribution.

NIREP NA0 database has already been released and distributed through XNAT central. NA0 can be downloaded at http://central.xnat.org/. Here, we have built up a project named "Non-rigid Image Registration Evaluation Project (NIREP)". People who want NA0 database for their research just need to register at XNAT central and request to get access to the NIREP project. After receiving their request, we add them as collaborators of NIREP project. That means, they can read, download and use project, subject and experiment data, but cannot edit or delete any content. Under XNAT NIREP project, people can find: (1) Image data. There are 16 MR sessions under the project, which refer to 16 datasets of NIREP NA0 database; (2) Description of the image data. At the top of the NIREP project page, a short introduction of NIREP can be found, as well as which images are included in each dataset; (3) Demographics of each subject. For each subject, its demographics can be found, such as age, gender, hand, race etc. Also under each MR session, more detail information of the MR images of each subject can be found too.

Figure 2.14 shows the XNAT central website and steps to download NA0 database. Dozens of researchers from around the world have downloaded NA0 for their research. XNAT central has been proved to be a useful way to release our database to general community.

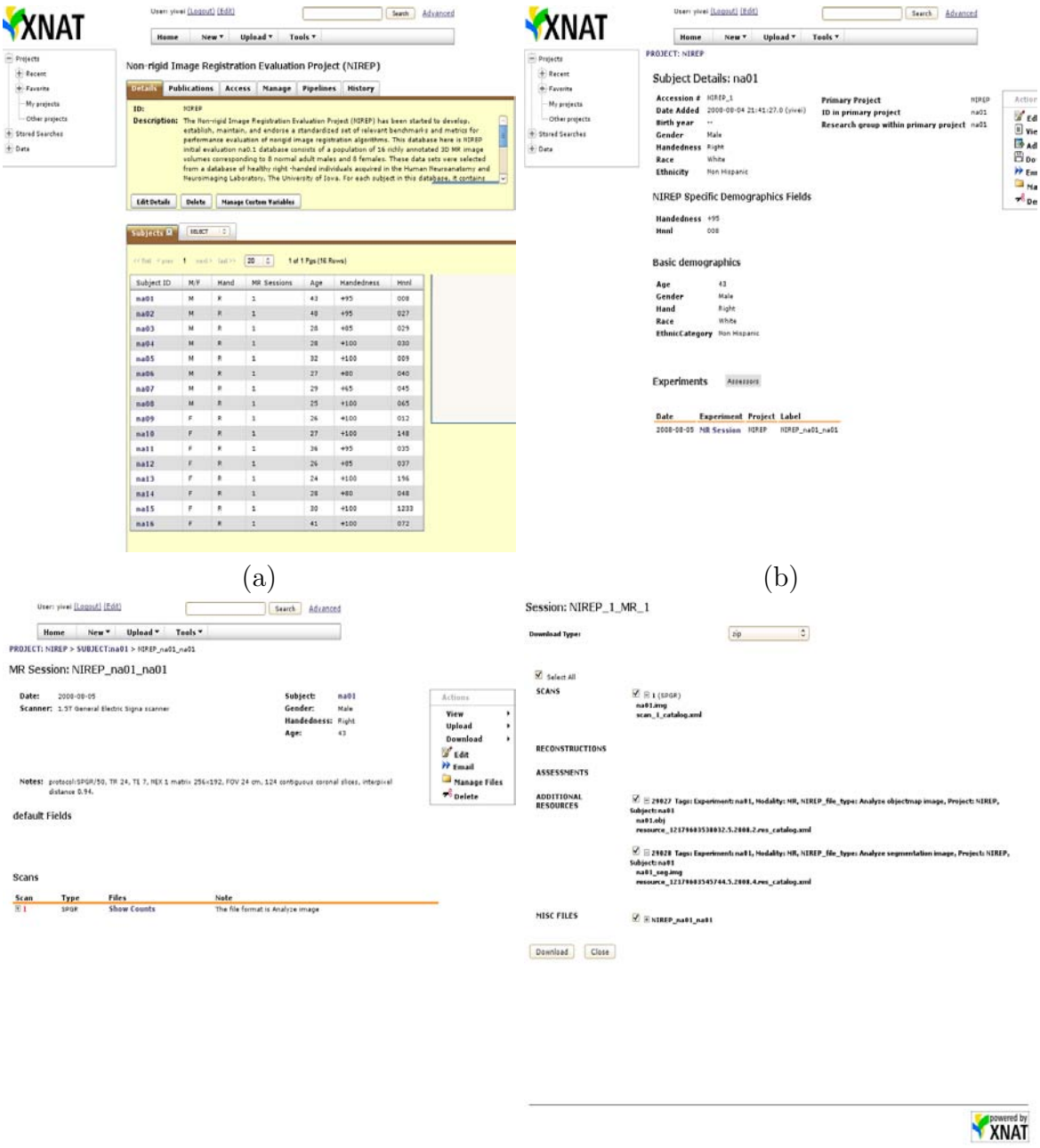


Figure 2.14: Snapshots of XNAT central and steps to download NA0 database (a)In the "project" table, click "NIREP" project. There are 16 MR sessions under the project. (b) Clicking on any of the subjects will allow people to see its MR session. (c) Click that MR session, MR images for that subject can be found. (d) To download the images, click "download images". Then there will be a download page, including "Scans" and "Additional resources".

2.3 Image Registration Methods

2.3.1 Introduction to Image Registration

Image registration has many uses in medical fields such as multi-modality fusion, image segmentation, deformable atlas registration, functional brain mapping, image guided surgery, and measuring growth and modeling motion. An image can be considered as a function taking from its domain, usually two-dimensional (2D) or three-dimensional (3D), which maps points in its coordinate system to intensity values, $G = [0, 255]$, for instant. These images are collected from imaging devices such as CT, MRI, cryosectioning, etc. Image registration is the process of geometrically aligning of two or more anatomical images and defining a meaningful correspondence mapping between them. The result transformation of the registration, which in general is a deformation field, is used to project data from one subject to another.

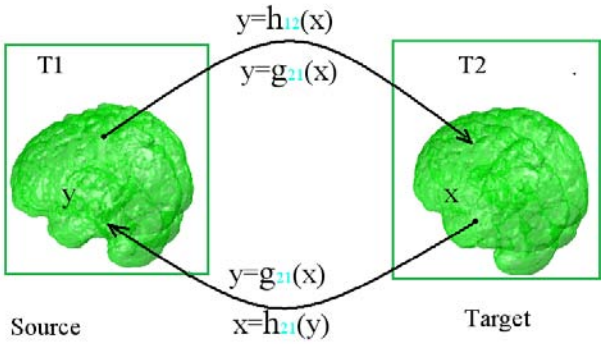


Figure 2.15: An illustration of the notation used to describe the image registration problem

Figure 2.15 illustrates some notation used to describe the image registration

problem. Given two images called *source* $T1$ and *target* $T2$, the goal of image registration is to find the optimal correspondence mapping or transformation which deforms $T1$ to $T2$. Let y denotes voxels in the coordinate system of $T1$ and x denotes voxels in the coordinate system of $T2$. We assume that both images are defined on the domain $\Omega \subset R^3$. The point-wise transformation from one image to the other can be represented by both Eulerian and Lagrangian reference systems. In an Eulerian reference system, a transformation defines a mapping from the initial location y to the destination location x as a function of the destination location x. In the a Lagrangian reference system, a transformation defines the mapping from initial location y to the destination location x as a function of the initial location y. The notation h is used to denote an Eulerian transformation and g is used to denote a Lagrangian transformation. The Eulerian transformation $h_{12}\colon \Omega \rightarrow \Omega$ maps a voxel $x \in \Omega$ to a voxel $y \in \Omega$ and $h_{21}\colon \Omega \rightarrow \Omega$ maps a voxel $y \in \Omega$ to a voxel $x \in \Omega$. The Lagrangian transformations $g_{12}\colon \Omega \rightarrow \Omega$ and $g_{21}\colon \Omega \rightarrow \Omega$ are related to the Eulerian transformations by the relationships $h_{12} = g_{21}$ and $h_{21} = g_{12}$, respectively.

There are different similarity functions used for image registration, like minimizing the relative overlap, maximizing the mutual information between two images, minimizing the distance between corresponding points of interest called *landmarks*, etc. All the similarity functions are designed to find the best correspondence mapping to match image $T1$ to $T2$.

2.3.2 Affine Registration

An affine transformation is any transformation that preserves co-linearity (i.e., all points lying on a line initially still lie on a line after transformation) and ratios of distances (e.g., the midpoint of a line segment remains the midpoint after transformation) [15]. Translation, rotation, rescaling, skewing are all affine transformations, as well as their combinations. To represent affine transformations with matrices, we must use homogeneous coordinates. This means representing a 3-vector (x, y, z) as a 4-vector (x, y, z, 1). The affine transformation, represented by a homogenous transformation equation for mapping the three dimensional coordinate (x,y,z) in target file to the three-dimensional coordinate (x', y', z') in template file is:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & m \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{2.1}$$

where a, b, c, d, e, f, g, h, i, j, k and m are independent parameters. This Affine registration model is called 3D affine 12 parameter model.

The Automated Image Registration (AIR 5.2.5) software was used to perform affine registration. This software can be downloaded at http://bishopw.loni.ucla.edu/AIR5/.

2.3.3 Automatic Image Registration

Automated Image Registration (AIR) [38] was developed by Roger Woods at the David Geffen School of Medicine at UCLA, CA. AIR is performed using both

linear and nonlinear models to register 3D (and 2D) images within and across subjects and within and sometimes across imaging modalities. AIR registers the source image to target image by minimizing the mean squared difference between these two images. Registration begins with the first order polynomial (i.e., 3D affine transformation) and the order is incremented until the order specified by the user is reached. Usually the final nonlinear polynomial spatial transformation is 3D fifth order nonlinear 168 parameters model. All transformations are stored as fifth order polynomials. For mapping the three dimensional coordinate (x, y, z) in target file to the three-dimensional coordinate (x', y', z') in template file, the array e computed by AIR is:

$$
\begin{aligned}
x' &= e[0][0] + e[1][0]\,x + e[2][0]\,y + e[3][0]\,z + e[4][0] * x^2 + \cdots + \\
&\qquad e[10][0] * x^3 + \cdots + e[35][0] * x^5 + \cdots + e[55][0] * z^5, \\
y' &= e[0][1] + e[1][1]\,x + e[2][1]\,y + e[3][1]\,z + e[4][1] * x^2 + \cdots + \\
&\qquad e[10][1] * x^3 + \cdots + e[35][1] * x^5 + \cdots + e[55][1] * z^5, \\
z' &= e[0][2] + e[1][2]\,x + e[2][2]\,y + e[3][2]\,z + e[4][2] * x^2 + \cdots + \\
&\qquad e[10][2] * x^3 + \cdots + e[35][2] * x^5 + \cdots + e[55][2] * z^5.
\end{aligned}
\tag{2.2}
$$

The AIR5.2.5 software package was used to register images with a 5^{th} order polynomial registration.

2.3.4 Demons Registration

Thirion's Demons algorithm [33, 34] is another common method used to deformably register two images to produce a non-rigid spatial transformation [35]. This method alternates between the computation of warping forces inspired from optical flow theory and the regularization of these forces by a simple Gaussian smoothing. The Insight Toolkit (ITK) provides an implementation of Thirion's "demons" algorithm [27], which is implemented as part of the finite difference solver (FDS) framework. In this implementation, each image is viewed as a set of iso-intensity contours. The main idea is that a regular grid of forces deforms an image by pushing the contours in the normal direction. The displacement or optical flow D(X) between the fixed image (i.e., target image) f(X) and the moving image (i.e., source image) m(X) is:

$$D(X) = -\frac{(m(X) - f(X))\nabla f(X)}{\|\nabla f\|^2 + (m(X) - f(X))^2/K} \tag{2.3}$$

where K is a normalization factor that accounts for the units imbalance between intensities and gradients. Starting with an initial deformation field $D^0(X)$, the demons algorithm iteratively updates the fields using Equation 2.3. An option to solve for the field uniquely is to enforce an elastic-like behavior, smoothing the deformation field with a Gaussian filter between iterations.

To improve the speed, registration using multi-resolution is used here. The basic idea is that registration is first performed at a coarse scale where the images have fewer pixels. The spatial mapping determined at the coarse level is then used to

initialize registration at the next finer scale. This process is repeated until it reaches the finest possible scale [8]. ITK offers a multi-resolution registration framework that is directly compatible with all the registration framework components.

Demons registration with multi-resolution was applied using the ITK software as discussed above.

2.3.5 SICLE

Small deformation inverse consistent linear elastic (SICLE) image registration algorithm was developed by Gary E. Christensen et. al [10, 11, 12, 25]. SICLE is a registration algorithm that jointly estimates the forward and reverse transformations between two images while minimizing the inverse consistency error between these transformations. Suppose S and T represent the source and the target images which are three-dimensional image volumes [7], and Ω_d is the discrete voxel lattice coordinates and let Ω_c be the corresponding continuous domain. Let h and g denote the forward transformation and reverse transformation respectively. SICLE image registration algorithm can be defined by minimizing the cost function:

$$\begin{aligned} C = \sigma \int_{\Omega_c} |S_c(h(x)) - T_c(x)|^2 + |T_c(g(x)) - S_c(x)|^2 dx \\ + \chi \int_{\Omega_c} ||h_c(x) - g_c^{-1}(x)||^2 + ||g_c(x) - h_c^{-1}(x)||^2 dx \\ + \rho \int_{\Omega_c} ||L_c u_c(x)||^2 dx + \int_{\Omega} ||L_c w_c(x)||^2 dx \end{aligned} \tag{2.4}$$

where the vector-valued functions u, w are called displacement fields since they define the transformation in terms of a displacement from a location x, L_c is the linear

elasticity operator, and the constants σ, χ and ρ are used to balance the constraints. The first integral of the cost function defines the cumulative intensity squared error (shape differences) between the deformed image $S \circ h$ and target image T and the differences between the deformed image $T \circ g$ and source image S. The second integral defines the inverse consistency error and is minimized when $h = g^{-1}$. This constraint couples the estimation of h and g together and penalizes transformations that are not inverses of each other, i.e., minimizing the inverse consistency error described in following section. The third intergral is a linear elastic constraint [13, 14] to ensure that the transformations maintain the topology of the images T and S. This term is used to regularize the forward and reverse displacement fileds so that they are smooth and continuous by penalizing large derivatives of the displacement fields. A 3D Fourier series representation [3] is used to parameterize the forward and reverse transformations. The Fourier series parameterization used in the consistent registration algorithm is described by Christensen and Johnson [11, 24] in detail. Multi-resolution method descibed in last section is taken in both frequency and spatial domains to estimate the Fourier Series coefficients by first estimating the low frequency components and then increasing the number of harmonics [26].

SICLE registration is applied by the program developed by Gary E. Christensen and others at the University of Iowa.

2.3.6 SLE

Small deformation linear elastic (SLE) image registration algorithm corresponds to SICLE image registration algorithm when $\sigma \neq 0$, $\rho \neq 0$ and $\chi = 0$.

That means there is no inverse consistent constraint in SLE cost function, and its cost function is equal to Equation 2.4 without the second integral.

SLE image registration is also applied by the program developed by Gary E. Christensen and others at the University of Iowa.

2.4 Registration Evaluation Methods

2.4.1 Average Relative Overlap

The alignment of objects, structures, organs, regions of interests (ROIs) is a good indicator of how well two images are registered. The relative overlap (RO) of segmentations is a measure of how well two corresponding segmented regions agree with each other. Here RO is computed for each ROI and this can measure how well two corresponding ROIs agree with each other. The RO for the kth ROI of both Image I_i and Image I_j registered to Image I_i is given by:

$$RO_{ij}^{k} = \frac{\sum_{x=1}^{M} ROI_i^k(x) \cap ROI_j^k(h_{ji}(x))}{\sum_{x=1}^{M} ROI_i^k(x) \cup ROI_j^k(h_{ji}(x))}, \tag{2.5}$$

where h_{ji} is the transformation to register Image I_j to Image I_i, $ROI_i^k(x)$ returns 1 if x is in the kth ROI in I_i, and returns 0 if is not, and M is the number of voxels. Thus, $\sum_{x=1}^{M} ROI_i^k(x) \cap ROI_j^k(h_{ji}(x))$ defines the number of voxels in the intersection of two corresponding ROIs and $\sum_{x=1}^{M} ROI_i^k(x) \cap ROI_j^k(h_{ji}(x))$ is the number of voxels in the union.

Then ARO of the kth ROI is defined as:

$$ARO^k = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{N-1}\sum_{j=1,j\neq i}^{N}\frac{\sum_{x=1}^{M} ROI_i^k(x) \cap ROI_j^k(h_{ji}(x))}{\sum_{x=1}^{M} ROI_i^k(x) \cup ROI_j^k(h_{ji}(x))}. \tag{2.6}$$

where N is the total number of images in the database, and h_{ji} is still the transformation to register Image I_j to Image I_i.

2.4.2 Normalized ROI Overlap

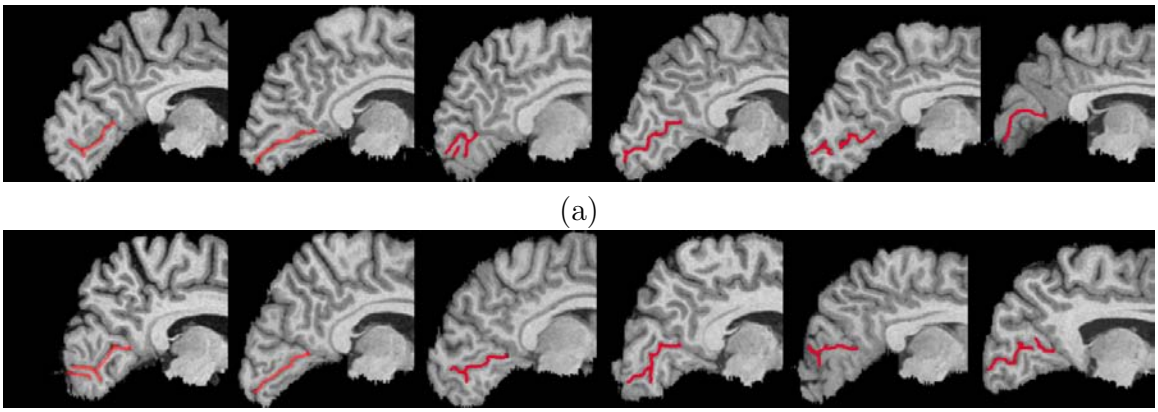
Normalized ROI overlap is the voxel-wise overlap generated by overlapping the ROIs from source image to the target image [19]. Suppose P_i is the target segmentation image, and the other $N-1$ source segmentation images P_j are deformed to P_i by transformation h_{ji}, the voxel-wise normalized ROI overlap to the target coordinate system is defined as:

$$RO_i(x) = \frac{1}{N-1}\sum_{j=1}^{N-1}\delta[P_j(h_{ji}(x)), P_i(x)] \tag{2.7}$$

where

$$\delta(x,y) = \begin{cases} 1 & if x = y \\ 0 & otherwise, \end{cases} \tag{2.8}$$

Visualization of the ROI overlap can easily show the number of corresponding ROIs among source images that agree with the ROI in target image.



(a)

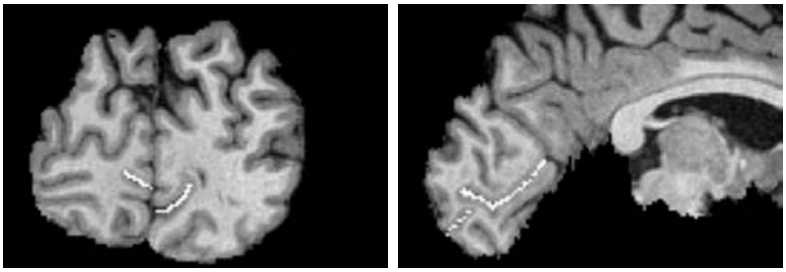
(b)

Figure 2.16: Calcarine Sulcus (a) left Calcarine sulcus of 6 datasets from NA1.(b)right Calcarine sulcus of 6 datasets from NA1.

2.4.3 Alignment of Calcarine Sulci

The calcarine sulcus is where the primary visual cortex is concentrated. As visual functional response is involved in our study, the alignment of calcarine sulci is a useful tool to show how the registration algorithm works and the evaluation result gives a good comparison baseline with evaluation by fMRI. In this study, the posterior portion of the calcarine sulcus was concerned. The posterior portion of the calcarine sulcus begins near the occipital pole and runs forward to a point a little below the splenium of the corpus callosum where it is joined at an acute angle by the medial part of the parietooccipital fissure. Due to complexity of the shape of the calcarine sulcus, it is a hard task to say how the pair-wise corresponding relationship between two calcarine sulci should be. Figure 2.16 shows the variability of calcarine sulci of 6 data sets of NA1 database.

To evaluate the alignment of calcarine sulci between target and source images,



(a) (b)

Figure 2.17: The pointset of Calcarine Sulcus shown in MRI of NA1. The White line here is refer to Calcarine Sulcus.(a) cornal. (b) sagital.

the calcarine sulci were first outlined as point sets. The Calcarine Sulci point sets (in Figure 2.17) of all the data sets in NA1 were created by master student Elizabeth K Broeder under the supervision of Joel Bruss. Then transformations calculated by five non-rigid registration algorithms were applied to these point sets and we got the deformed point sets.

Finally, Hausdorff distance was used for evaluation. Hausdorff distance measures how far two non-empty subsets are from each other. Here, we define average Hausdorff distance of two calcarine sulcus point sets A and B as:

$$H(A,B) = \frac{1}{2}(h(A,B) + h(B,A)) \tag{2.9}$$

where

$$h(A,B) = \max_{a \in A} \min_{b \in B} \|a - b\| \tag{2.10}$$

Informally, two sets are close in the Hausdorff distance if every point of either set is close to some point of the other set. So that the algorithm with minimal Hausdorff

distance between deformed calcarine sulcus and target one, can be considered the best algorithm with respect to this statistic evaluation.

2.4.4 Intensity Variance

A common method used to measure image registration performance is to register a population of images with a target image and average the intensities of the registered images. The idea is that the better the registration algorithm is, the closer each registered image looks to the target image and the sharper the intensity average image. One way to measure the sharpness of the intensity average image is to compute the variance of the registered intensity images.

The voxel-wise intensity variance (IV) of a population of M images registered to image j is computed as:

$$IV_j(x) = \frac{1}{M-1}\sum_{i=1}^{M}(T_i(h_{ij}(x)) - ave_j(x))^2 \tag{2.11}$$

where

$$ave_j(x) = \frac{1}{M}\sum_{i=1}^{M}T_i(h_{ij}(x)) \tag{2.12}$$

T_i is the ith image of the population and $h_{ij}(x)$ is the transformation from image i to j with respect to a Eulerian coordinate system.

2.4.5 Inverse Consistensy Error

Inverse consistency and transitivity consistency evaluate registration performance based on desired transformation properties [11, 12, 25].In image registration,

it is often assumed that the correspondence mapping between two anatomical images is one-to-one, i.e., each point in source image S is mapped to only one point in target image T and vice versa. However, the forward mapping from S to T and the reverse mapping from T to S resulted from many image registration techniques are not inverses of each other. This inconsistency is a result from the inability to uniquely describe correspondence between two images in the registration process [7]. Figure 2.18 illustrates the inverse consistency error (ICE) involved in the image registration. The ICE is the difference between $h_{12}(h_{21}(y))$ and y, i.e., transform a point from brain 1 to brain 2, then from brain 2 to brain 1 should return to the original point.

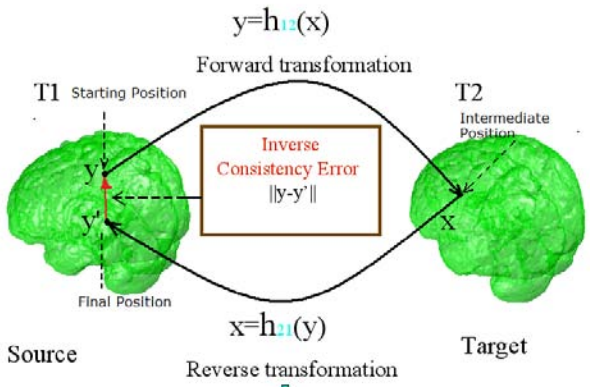


Figure 2.18: An illustration of the mapping ambiguity problem and the inverse consistency error in the image registration of two images.

As shown in Figure 2.18, the inverse consistency error statistic measures the error between a forward and reverse transformation between two images. Ideally, the forward transformation equals the inverse of the reverse transformation implying

a consistent definition of correspondence between two images, i.e., correspondence defined by the forward transformation should be the same as that defined by the reverse transformations. Thus, composing the forward and reverse transformations together produces the identity map when there is no inverse consistency error. The inverse consistency error is defined as the difference between the composition of the forward and reverse transformations and the identity mapping.

The voxel-wise inverse consistency error with respect to source image j is computed as

$$ICE_j(y) = \frac{1}{M}\sum_{i=1}^{M} ||h_{ji}(h_{ij}(y)) - y|| \tag{2.13}$$

where h_{21} is the transformation from image 2 to 1, M is the number of images in the evaluation population and $||\cdot||$ is the standard Euclidean norm. And ICE can be represented as an image, that map points in source coordinate system to ICE values.

Another way to calculate inverse consistency error which is mapped in target coordinate system is defined as:

$$ICE_i(x) = \frac{1}{M}\sum_{j=1}^{M} ||h_{ji}(x) - h_{ij}^{-1}(x)|| \tag{2.14}$$

where h_{ij}^{-1} is the inverse of transformation h_{ij} which maps x in target image to y in source image, M is the number of images in the evaluation population and $||\cdot||$ is the standard Euclidean norm. ICE defined by Equation 2.13 was applied in this work, as the computation of inverse of transformation (as in Equation 2.14) is time consuming.

2.4.6 Transitivity Error

The transitivity error (TE) statistic evaluates how well pair-wise registrations of the image population satisfy the transitivity property [11, 12, 25]. The transitivity property is important to minimize inconsistency error when two transformations are composed together to represent one transformation. Figure 2.19 illustrates the transitivity error involved in image registration. The transitivity error is the difference between $h_{12}(h_{23}(h_{31}(y))$ and y, i.e., transforming a point from brain 1 to 3, then from 3 to 2, and then from 2 to 1 should return to the original point. Any difference between the starting point and ending point is defined as the transitivity error.

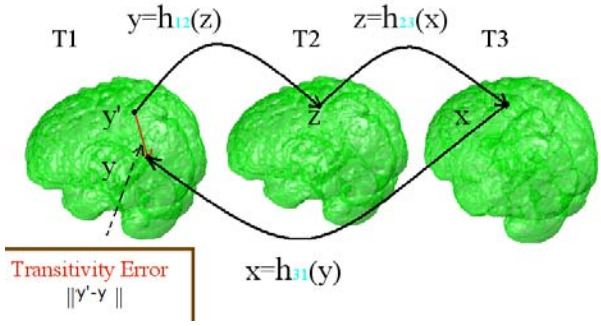


Figure 2.19: An illustration of the transitivity error in the image registration of two images.

Ideally, transformations that define correspondences among three images should project a point from image $T1$ to $T2$ to $T3$, and to $T1$ back to the original position. The transitivity error for a set of transformations is defined as the error difference between the composition of the transformations between three images and the identity

map.

The voxel-wise transitivity error with respect to source image i is computed as:

$$TE_i(y) = \frac{1}{(M-1)(M-2)} \sum_{\substack{j=1 \\ j \neq i}}^{M} \sum_{\substack{k=1 \\ k \neq j \\ k \neq i}}^{M} ||h_{ij}(h_{jk}(h_{ki}(x))) - x||. \tag{2.15}$$

where i is the image $T1$ shown is Figure 2.19, j is the image $T2$ shown in Figure 2.19, k is the image $T3$ shown in Figure 2.19, M is the number of images in the evaluation population and $|| \cdot ||$ is the standard Euclidean norm.

TE defined in Equation 2.15 requires a lot of computations because all combinations of pair-wise transformations defining correspondences in a population must be composed together. We used another way to calculate transitivity error that requires less computation in this work. It is defined as:

$$TE_k(x) = \frac{1}{(M-1)(M-2)} \sum_{\substack{i=1 \\ i \neq k}}^{M} \sum_{\substack{j=1 \\ j \neq i \\ j \neq k}}^{M} ||(h_{ij}(h_{jk}(x)) - h_{ik}(x)||. \tag{2.16}$$

Because this equation just requires transform the transformation once, whereas Equation 2.15 does twice. We plan to investigate ways to reduce this computation burden such as using smaller population sizes and computing the TE at a small number of randomly selected points in the image domain.

2.4.7 Discussion of registration evaluation by fMRI

2.4.7.1 MRI with functional markers for registration evaluation

The cerebral cortex is the largest graymatter structure in the nervous system and widely agreed to be differentiated into functional components supporting key aspects of sensory, motor, and higher function.

We proposed here an approach to evaluate registration algorithms by exploiting functional properties of visual-related cortex in the occipital lobes. The primary visual cortex bears a close relationship to the calcarine sulcus, which is present in the mesial occipital lobe of all human brains. Certain basic structural-functional relationships are reliable: the right visual hemifield maps to the left calcarine region, and the left hemifield to the right calcarine region; the horizontal meridia are represented in the fundi of the calcarine sulci; the superior quadrants are represented in the lower banks of the calcarine sulci and adjacent inferocalcarine occipital lobe, and the inferior quadrants on the superior bank and supracalcarine occipital lobe. The eccentricity axis of the visual field also has a principled relationship to anatomy: central vision is represented posteriorly, on the occipital pole, while the peripheral fields are represented anteriorly. Secondary visual areas (V2, V3, etc.) are also retinotopically mapped. The visual field "sign", or polarity, of the visual field representation is either similar to the actual visual field geometry or mirror symmetrical to it. The field sign typically reverses between adjacent visual areas, and the line of reversal is therefore a functional fiducial for areal boundaries [31].

Figure 2.20 shows an example of the visual stimulation fMRI and the retino-

topic maps of visual areas with the corresponding eccentricity and polar-angle axis.

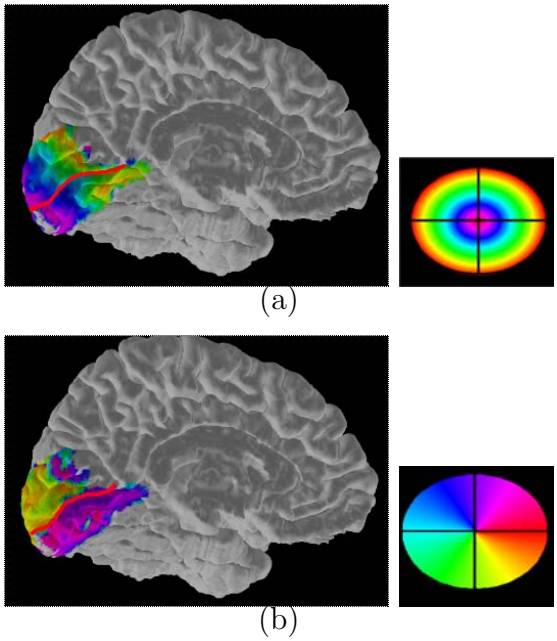


Figure 2.20: Polar coordinate retinotopic mapping with fMRI, polar angle and eccentricity phase maps are presente in which the voxel color corresponds to the polar angle wedges and eccentricity annuli shown in the keys on the right. (a) eccentricity (b) polar angle

We use these reliable aspects of structural-functional correspondences to evaluate whether a particular anatomical transformation improves the functional correspondence between two brains, and in turn to evaluate the performance of registration algorithms. In this work, 6 datasets from NA1 were involved and they were: 3065, 3362, 3368, 3402, 3407, 3413. T1 images of these 6 datasets were registered by different registration algorithms and their performance were compared by mapping

segmentations and fMRI through computed transformations. The functional images (as shown in Figure 2.20) used here represented phase value of response under eccentricity stimulation, in the whole occipital region, i.e., response in graymatter as well as whitematter and csf.

2.4.7.2 Evaluation method 1: Normalized Cross Correlation in whole occipital region

Normalized Cross Correlation (NCC) is calculated in the intersection area (as shown in Figure 2.21) of two fMRIs in the whole occipital region.

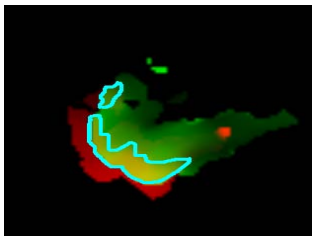


Figure 2.21: Intersection area (the yellow region within the blue curves) of two functional response (one is shown as red area and the other is green).

NCC is defined as:

$$NCC_{ij} = \frac{1}{n-1} \sum_{x \in \Omega_k} \frac{(F_i(x) - \bar{F}_i)(F_j(x) - \bar{F}_j)}{\sigma_i \sigma_j} \tag{2.17a}$$

$$\sigma_i = \sqrt{\frac{1}{n-1} \sum_{x \in \Omega_k} (F_i(x) - \bar{F}_i)^2} \tag{2.17b}$$

$$\bar{F}_i = \frac{1}{n} \sum_{x \in \Omega_k} F_i(x) \tag{2.17c}$$

where Ω_k is the intersection area of two fMRIs F_i and F_j , and n is the total number

of voxels in Ω_k.

If F_j is deformed to target F_i, NCC of F_i and deformed fMRI F_{ji} should be:

$$NCC_{ij} = \frac{1}{n-1} \sum_{x \in \Omega_k} \frac{(F_i(x) - \bar{F}_i)(F_j(h_{ji}(x)) - \bar{F}_{ji})}{\sigma_i \sigma_{ji}} \tag{2.18a}$$

$$\sigma_{ji} = \sqrt{\frac{1}{n-1} \sum_{x \in \Omega_k} (F_j(h_{ji}(x)) - \bar{F}_{ji})^2} \tag{2.18b}$$

$$\bar{F}_{ji} = \frac{1}{n} \sum_{x \in \Omega_k} F_j(h_{ji}(x)) \tag{2.18c}$$

where h_{ji} is the computed transformation to register F_j to target F_i to get the deformed functional image F_{ji}, and Ω_k is the intersection area of two functional MRI F_i and F_{ji} , and n is the total number of voxels in Ω_k.

6 datasets were included so far, so that there were 6 fMRIs and 30 pair-wise registration results. Averaged NCC (ANCC) is defined as:

$$ANCC = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{N-1} \sum_{j=1, j \neq i}^{N} NCC_{ij} \tag{2.19}$$

where N is equal to 6 in this work.

2.4.7.3 Evaluation method 2: Normalized Cross Correlation in Region of Interests (ROIs)

Furthermore, NCC was calculated in different ROIs separately. As functional response here just occurred in occipital region, just 8 ROIs were considered here. These 8 ROIs (4 in left brain and 4 in right brain) are:

Occipital region (as shown in Figure 2.22.)

Lcun: Left Cuneus;

Llocci: Left Lateral Occipital Gyrus;

Lling: Left Lingual Gyrus;

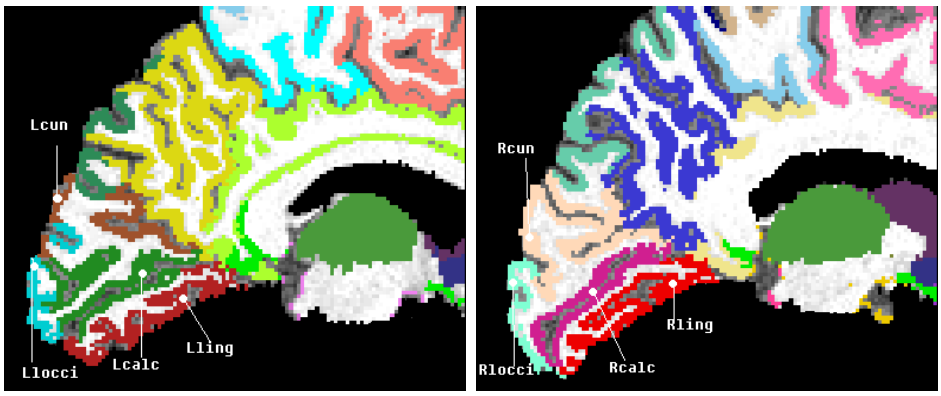
Lcalc: Left Calcarine (pericalcarine) Region;

Rcun: Right Cuneus;

Rlocci: Right Lateral Occipital Gyrus;

Rling: Right Lingual Gyrus;

Rcalc: Right Calcarine (pericalcarine) Region;



(a) (b)

Figure 2.22: Occipital Regions (a)left (b)right

NCC is calculated in the intersection area of two fMRIs within the intersection of two corresponding ROIs of two segmentation images (shown in Figure 2.23).

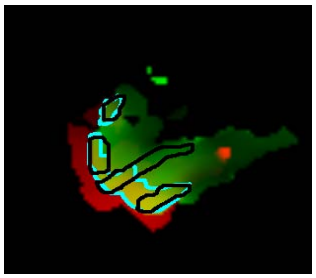


Figure 2.23: Statistic value is calculated in intersection area of two functional response (yellow region in blue curves) in each ROI (black curves).

NCC in ROI R is defined as:

$$NCC_{ij}^{R} = \frac{1}{n-1} \sum_{x \in \Omega_k^R} \frac{(F_i(x) - \bar{F_i^R})(F_j(x) - \bar{F_j^R})}{\sigma_i \sigma_j} \tag{2.20a}$$

$$\sigma_i = \sqrt{\frac{1}{n-1} \sum_{x \in \Omega_k^R} (F_i(x) - \bar{F_i^R})^2} \tag{2.20b}$$

$$\bar{F_i^R} = \frac{1}{n} \sum_{x \in \Omega_k^R} F_i(x) \tag{2.20c}$$

where Ω_k^R is the intersection area of two fMRIs F_i and F_j in ROI R, and n is the total number of voxels in Ω_k^R.

If F_j is deformed to target F_i, NCC of F_i and deformed functional image F_{ij}

in ROI R should be:

$$NCC_{ij}^{R} = \frac{1}{n-1} \sum_{x \in \Omega_k^R} \frac{(F_i(x) - \bar{F_i^R})(F_j(h_{ji}(x)) - \bar{F_{ji}^R})}{\sigma_i \sigma_{ji}} \tag{2.21a}$$

$$\sigma_{ji} = \sqrt{\frac{1}{n-1} \sum_{x \in \Omega_k^R} (F_j(h_{ji}(x)) - \bar{F_{ji}^R})^2} \tag{2.21b}$$

$$\bar{F_{ji}^R} = \frac{1}{n} \sum_{x \in \Omega_k^R} F_j(h_{ji}(x)) \tag{2.21c}$$

where h_{ji} is the computed transformation to register F_j to target F_i to get the deformed fMRI F_{ji}, and Ω_k is the intersection area of two fMRIs F_i and $F_j i$ in ROI R, and n is the total number of voxels in Ω_k.

2.5 STARD Documentation for NIREP work

The STARD (Standards for Reporting of Diagnostic Accuracy) statement was developed by a group of scientists and editors to improve the accuracy and completeness of reporting of studies of diagnostic accuracy. This statement consists of a checklist of 25 items and a flow diagram to ensure that all relevant information of a diagnostic accuracy study is present. It also helps readers to better understand diagnostic reports, assess their quality, assess the potential for bias in the study, and evaluate the generalization of the study [4, 5]. Our study follows the model of the STARD statement to disclose all relevant information for each non-rigid registration evaluation test, as one goal of NIREP is to accurately report the registration evaluation results so that other investigators can get access to all the details of the experiments preformed and they do not need to repeat the work. The STARD doc-

umentation in this work described each checklist item recommended in the STARD statement and gave the required documentation for the NA0 database of 16 MRIs with 32 ROIs and NA1 database of 18 MRIs with 57 ROIs and the evaluation statistics performance of different registration algorithms involved in this work. One big benefit of this documentation is that people who evaluate their own new registration algorithms with these two databases can refer back to this STARD documentation of NIREP and only have to change the items that are different to complete their own STARD documentation and then report to research community.

CHAPTER 3
RESULTS

Five non-rigid registration algorithms were evaluated using common databases. Rigid registration during preprocessing steps was used as the starting point for all non-rigid registration algorithms. It did not appear to be evaluated, yet was included to serve as a comparison baseline for some statistics in our study.

All 16 data sets of NA0 and 6 data sets of NA1 were involved in this work while other NA1 data sets will be involved in the future. This is to reduce the computation work and enough to show an example to evaluate performance of registration algorithms.

The statistical significance of the differences in the evaluation statistics were assessed using Student's t-test. For each statistic, the t-test was applied for each region of interest (32 ROIs for NA0 and 57 ROIs for NA1), for each two registration algorithms under each evaluation statistic, and for both NA0 and NA1 respectively. Alpha was set equal to 0.05. A two-tailed two-sample equal variance t-test was used for relative overlap and intensity variance. A two-tailed paired t-test was used for inverse consistency error and transitivity error. A significant difference between two registration algorithms was reported if the p value was less than 0.05 (Alpha).

Table 3.1: Averaged Relative Overlap of NA0

Registration Algorithms	Averaged ARO in 32 ROIs
Rigid	0.2643
Affine	0.3241
AIR	0.3578
Demons	0.5029
SLE	0.3891
SICLE	0.3814

3.1 Evaluation Results for ARO

3.1.1 NA0 database

The ARO performance for five non-rigid registration algorithms with respect to NA0 database are shown in Figure 3.1. ARO of rigid registration is also shown in this graph as a benchmark for non-rigid registration algorithms. The ARO statistic values were computed for each of the 32 ROI. As each pair-wise registration had a statistic value, the statistic calculation program sorted these statistic values in an ascending order. The data file recorded different percentile ranges for this order. Figure 3.1 shows the mean and the 05, 25, 75 and 95 percentile ranges for each algorithm. Each candlestick records different percentile ranges for this order. The values of ARO averaged over all of the 32 ROIs for each registration algorithm are shown in Table 3.1.2.

By applying Student's t-test, we found under the significant level 0.05: with respect to rigid registration, all 5 non-rigid registration algorithms provide significantly different, i.e., smaller ARO values of all the 32 regions than it; Also, 31 out of 32

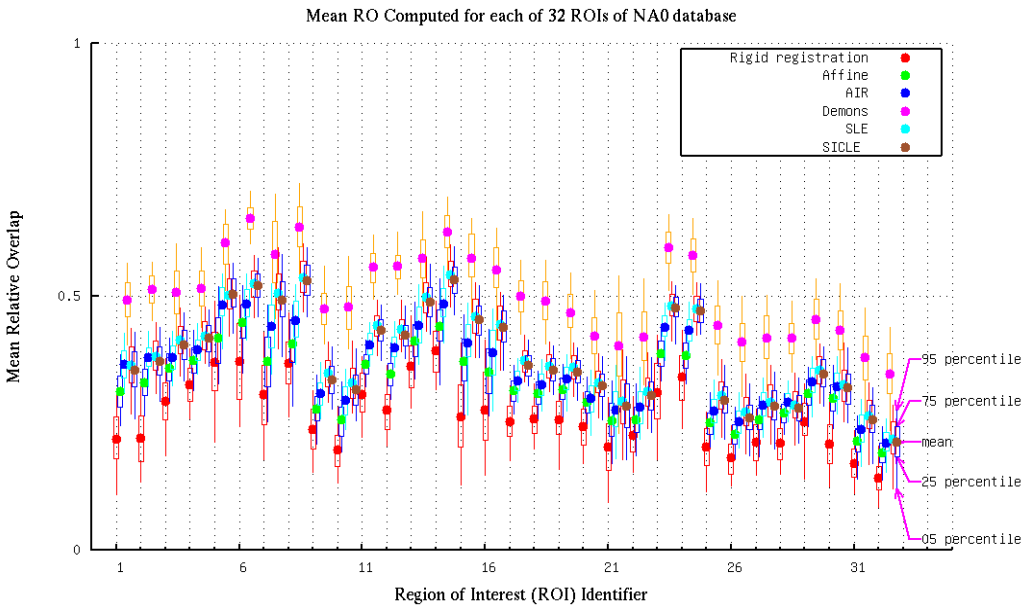


Figure 3.1: Graphs of average relative overlap (ARO) of 16 NA0 datasets, 240 transformations. Each measurement corresponds to the ARO computed with respect to a differnt template.

regions have significant difference between Affine and AIR using ARO; SLE, SICLE, and Demon have significantly different, i.e., smaller ARO over all the 32 regions than Affine; for AIR, 25 out of 32 regions have significant difference between it and SLE, and 23 out 32 between it and SICLE; the ARO for just 4 out of 32 have significant difference between SICLE and SLE; and Demons provides significantly different, i.e., smaller ARO of all the 32 regions than all the other registration algorithms.

3.1.2 NA1 database

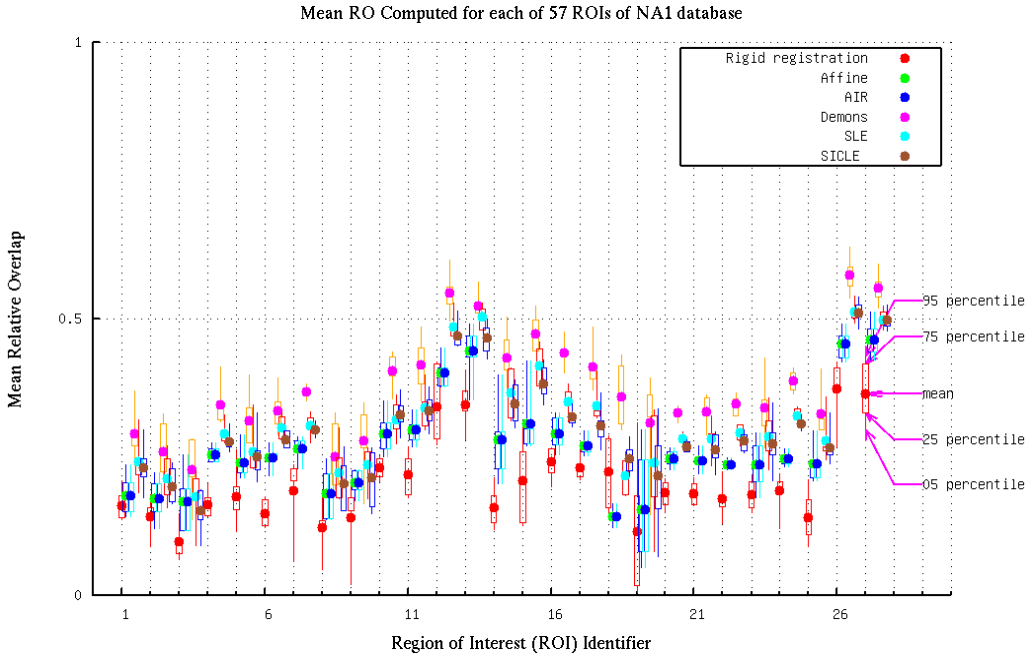
The ARO performance for five non-rigid registration algorithms with respect to NA1 database are shown in Figure 3.2. ARO of rigid registration is also shown in this graph as a benchmark for non-rigid registration algorithms. The ARO statistic

Table 3.2: Averaged Relative Overlap 0f NA1

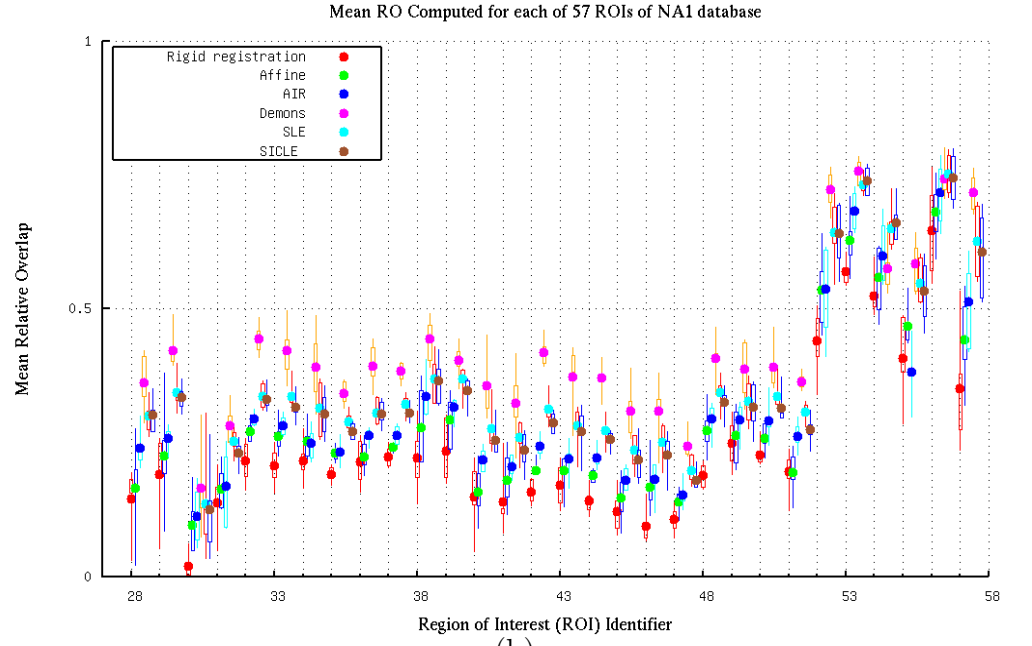
Registration Algorithms	Averaged ARO in 57 ROIs
RIgid	0.2202
Affine	0.2542
AIR	0.2874
Demons	0.4033
SLE	0.3441
SICLE	0.3302

values were computed for each of the 57 ROI. As each pair-wise registration had a statistic value, the statistic calculation program sorted these statistic values in an ascending order. The data file recorded different percentile ranges for this order. Figure 3.2 shows the mean and the 05, 25, 75 and 95 percentile ranges for each algorithm. Each candlestick records different percentile ranges for this order. The values of ARO averaged over all of the 57 ROIs for each registration algorithm are shown in Table 3.1.2.

By applying Student's t-test, we found under the significant level 0.05: with respect to rigid registration, 53 out of 57 regions have significant difference between it and Affine using ARO; all other 4 non-rigid registration algorithms provide significantly different, i.e., smaller ARO values of all the 57 regions than original; 56 out of 57 regions have significant difference between Affine and AIR using ARO; SLE, SICLE, Demons have significant difference, i.e., smaller ARO over all the 57 regions than Affine; for AIR, 49 out of 57 regions have significant difference between it and SLE, and 55 out of 57 between it and SICLE; the ARO for just 1 out of 57 have signifi-



(a)



(b)

Figure 3.2: Graphs of average relative overlap (ARO) of 6 NA1 datasets, 30 transformations. Each measurement corresponds to the ARO computed with respect to a differnt template. (a)ARO computed for Region of Interest Identifier from 1 to 27 (b)ARO computed for Region of Interest Identifier from 28 to 57.

cant difference between SICLE and SLE; and Demons provides significantly different, i.e., smaller ARO of all 57 regions than all the other registration algorithms.

3.2 Evaluation Results for Normalized ROI overlap

3.2.1 NA0 database

Figure 3.3 shows the voxel-wise ROI overlap generated by overlapping the ROIs from 15 other segmentation images of NA0 to the target segmentation image na01 of NA0. The reference coordinate system is the target's coordinate system. The other 15 images have similar ROIs that were mapped to the reference coordinate system. The numbers of corresponding ROIs that agree with the ROI of the target image after registration by different algorithms are shown in Figure 3.3 of NA0.

3.2.2 NA1 database

Figure 3.4 shows the voxel-wise ROI overlap generated by overlapping the ROIs from 5 other segmentation images of NA1 to the target segmentation image 3065 of NA1. The reference coordinate system is the target's coordinate system. The other 5 images have similar ROIs that were mapped to the reference coordinate system. The numbers of corresponding ROIs that agree with the ROI of the target image after registration by different algorithms are shown in Figure 3.4 of NA1.

3.3 Evaluation Results for Alignment of calcarine sulci

The transformations calculated by five non-rigid registration algorithms as well as rigid registration were applied to those 6 calcarine sulci point sets of NA1, and then the relative overlap (RO) statistic was calculated after registration by different

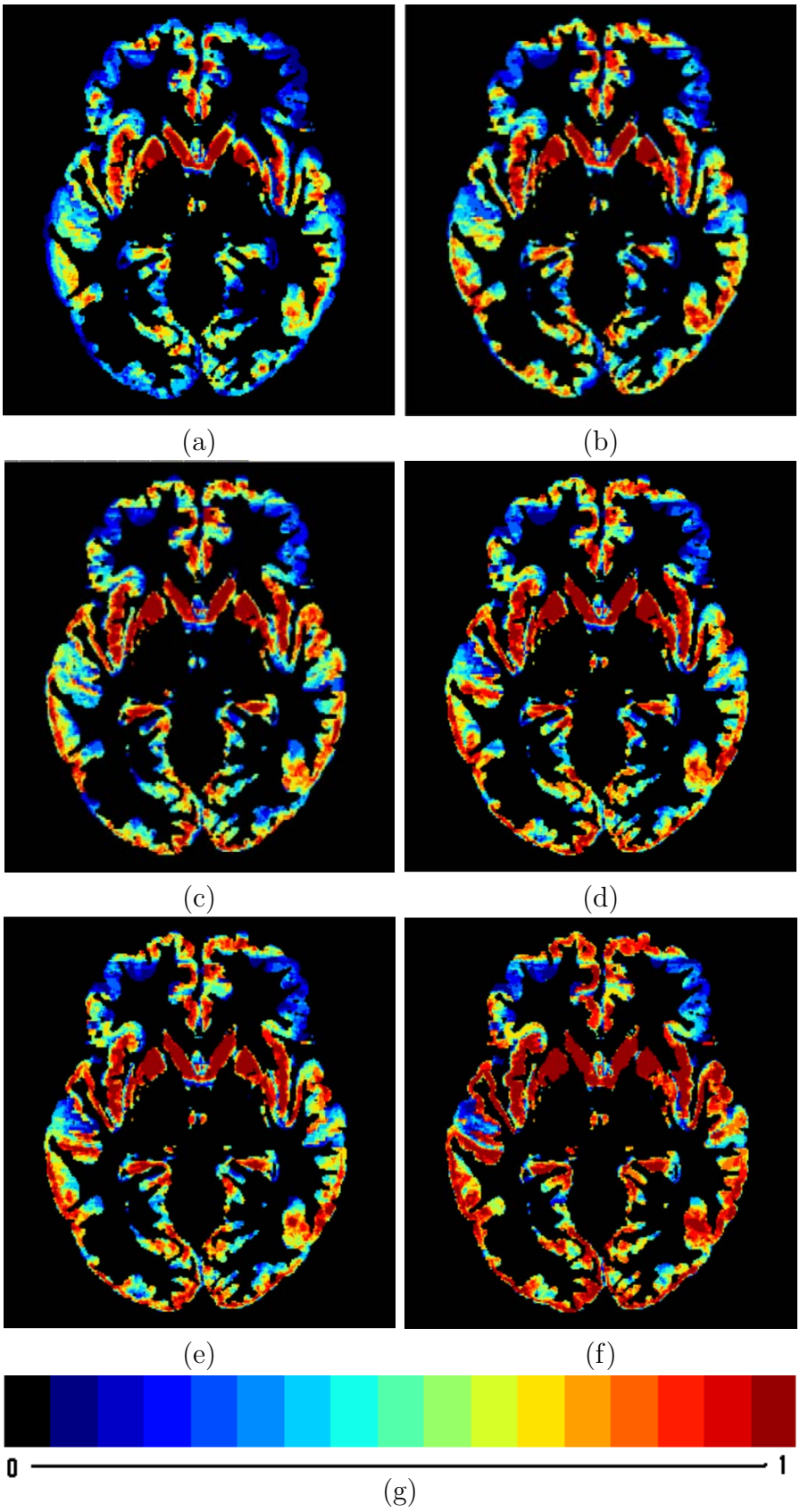


Figure 3.3: Normalized voxel ROI overlap of NA0 (a)Rigid (b)affine (c)AIR (d)SICLE (e)SLE (f)Demons (g)colorbar.The images are ordered from worst performance (left and top) to best performance (right and bottom).

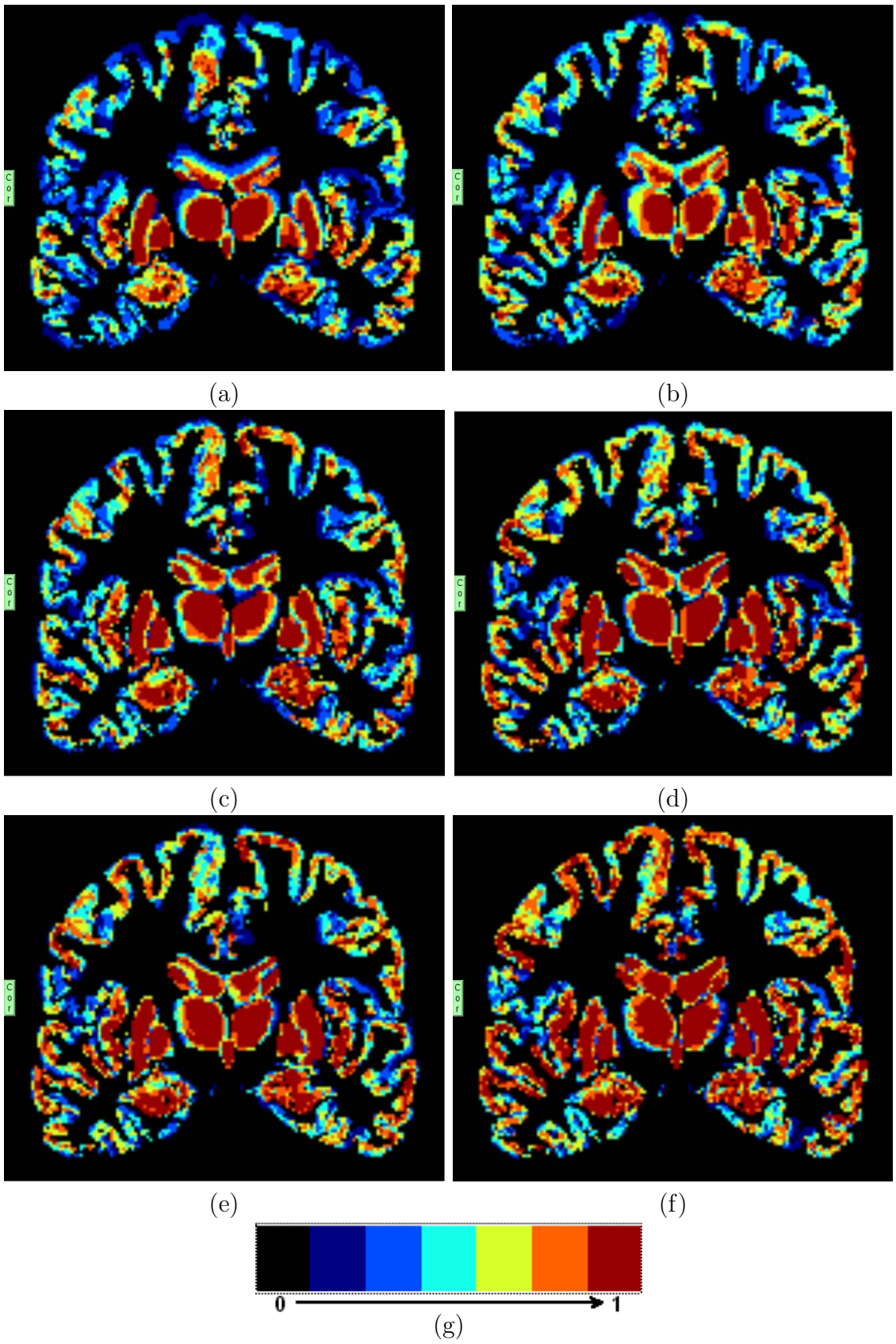


Figure 3.4: Normalized voxel RO of NA1 (a)Rigid (b)affine (c)AIR (d)SICLE (e)SLE (f)Demons (g)colorbar. The images are ordered from worst performance (left and top) to best performance (right and bottom)

Table 3.3: Average Hausdorff distance 0f NA1 database

Registration Algorithms	Averaged Hausdorff distance of left calcarine sulci	Averaged Hausdorff distance of right calcarine sulci
Rigid	5.25	6.90
Affine	4.403	5.717
AIR	3.734	4.302
Demons	3.037	3.376
SLE	3.372	4.06
SICLE	3.376	4.25

algorithms. Figure 3.5 shows the closer of calcarine sulci after non-rigid registration than rigid registration. All the slices in Figure 3.5 refer to of slice y=67 of subject 3065, which served as reference coordinate system here. Each color represents calcarine sulcus of one subject. And both left and right calcarine sulci are shown in these images. Yellow represents the target subject 3065.

Table 3.3 shows the average Hausdorff distance of 6 calcarine sulci point sets of NA1 database.

3.4 Evaluation Results for IV

3.4.1 NA0 database

The graph in Figure 3.6 plots the statistic value of the IV for each ROI among all the five non-rigid registration algorithms as well as original rigid registration. Table 3.4.1 shows the averaged value of IV over all 32 ROIs for each registration algorithm.

By applying Student's t-test, we found under the significant level 0.05: with respect to rigid registration, all other 5 non-rigid registration algorithms provide

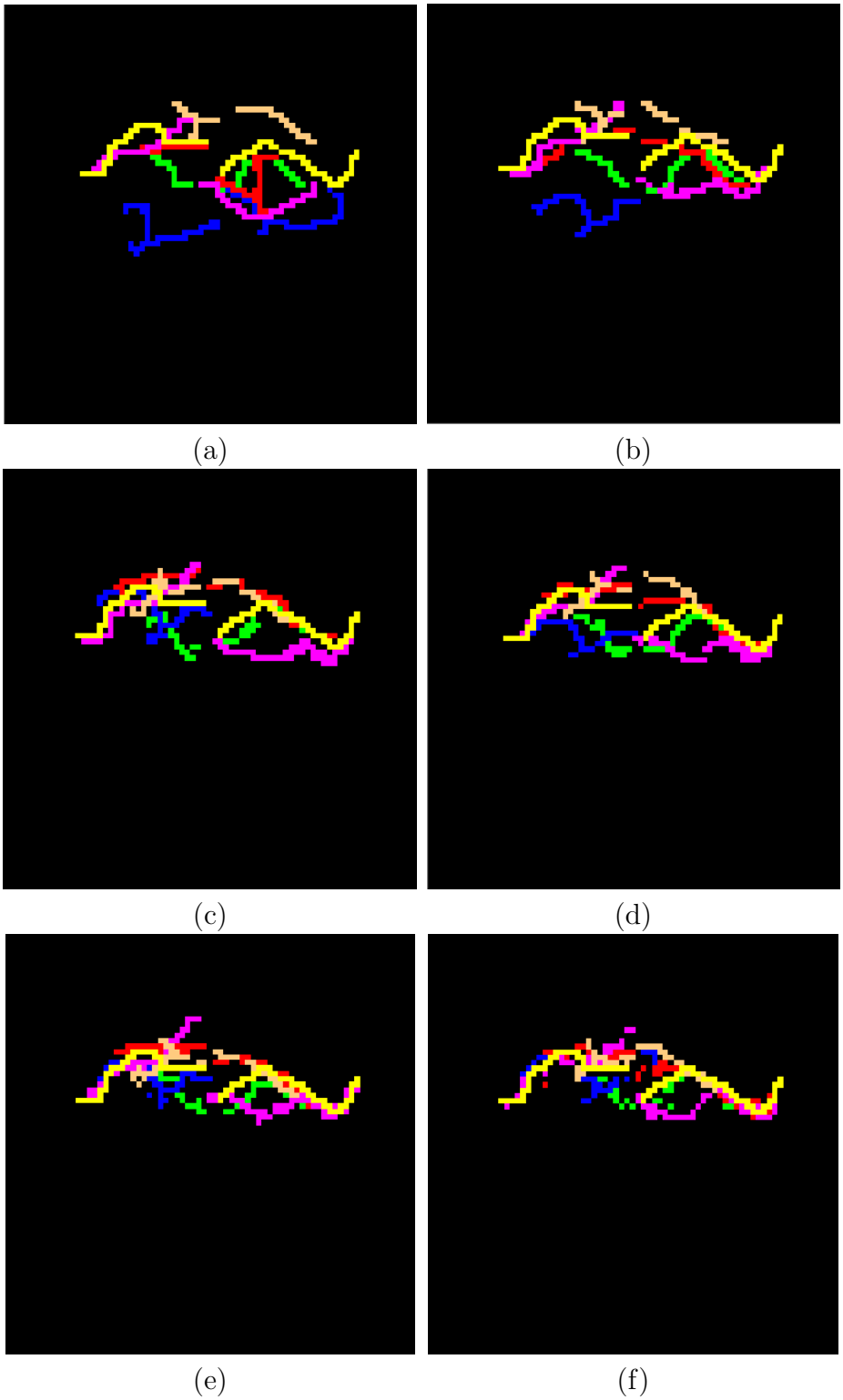


Figure 3.5: Alignment of Calcarine Sulcus (a)rigid (b)affine (c)AIR (d)SICLE (e)SLE (f)Demons. The images are ordered from worst performance (left and top) to best performance (right and bottom).

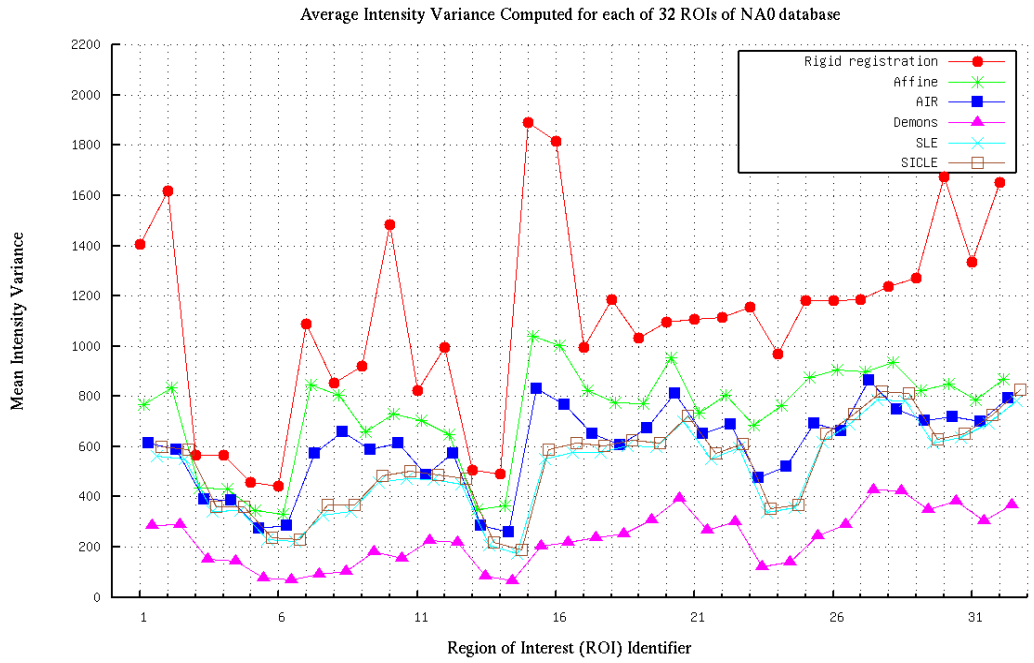


Figure 3.6: Graphs of average Intensity Variance (IV) of 16 NA0 datasets, 240 transformations. Each measurement corresponds to the IV computed with respect to a differnt template.

Table 3.4: Averaged Intensity Variance(IV) of NA0

Registration Algorithms	Average IV in 32 ROIs	IV standard diviation in 32 ROIs
Rigid	1103.14	169.06
Affine	738.59	97.95
AIR	598.33	70.66
Demons	230.24	43.88
SLE	505.81	47.31
SICLE	530.78	53.27

Table 3.5: Averaged Intensity Variance(IV) of NA1

Registration Algorithms	Average IV in 57 ROIs	IV stadard diviation in 57 ROIs
rigid	1540	367.593
Affine	1129.45	250.607
AIR	849.763	184.766
Demons	398.876	123.155
SLE	704.044	145.57
SICLE	722.688	148.785

significantly different, i.e., smaller IV values of all the 32 regions than it; 17 out of 32 regions have significant difference between Affine and AIR using IV; 31 out of 32 regions have significant difference between Affine and SICLE, and the same situation between SLE and Affine; Demons has significantly different, i.e., smaller IV over all the 32 regions than Affine; for AIR, 15 out of 32 regions have significant difference between it and SICLE, and 24 out of 32 between it and SLE; the IV for all 32 regions have no significant difference between SICLE and SLE; and Demons provides significantly different, i.e., smaller IV of all the 32 regions than all the other registration algorithms.

3.4.2 NA1 database

The graph in Figure 3.7 plots the statistic value of the IV for each ROI among all the five non-rigid registration algorithms as well as original rigid registration. Table 3.4.2 shows the averaged value of IV over all 57 ROIs for each registration algorithm.

By applying Student's t-test, we found under the significant level 0.05: with

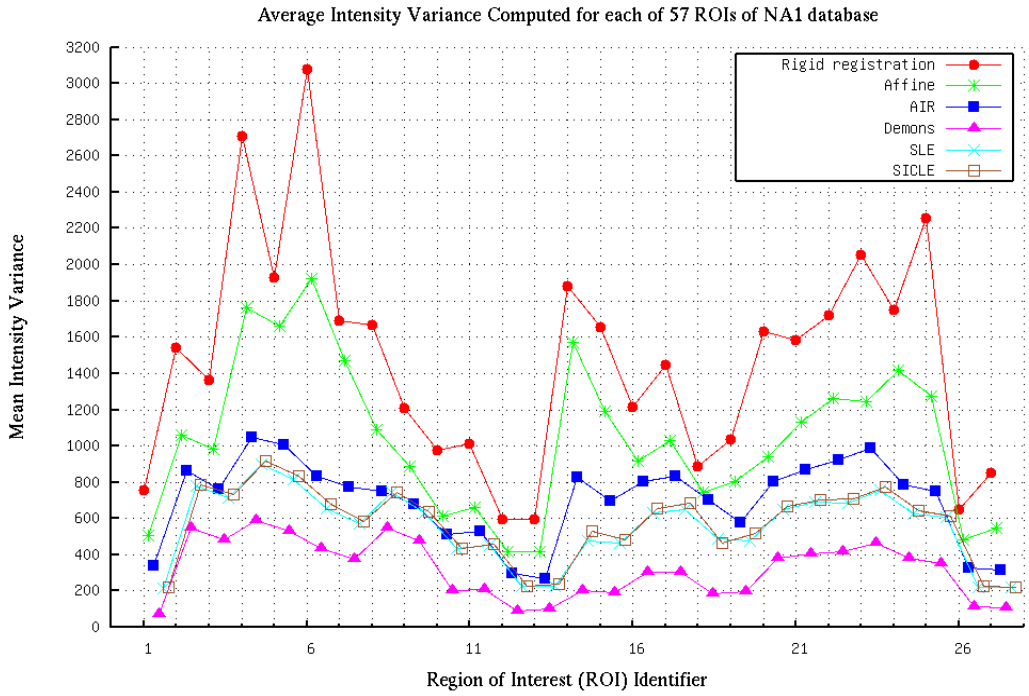


Figure 3.7: Graphs of average Intensity Variance (IV) of 6 NA1 datasets, 30 transformations. Each measurement corresponds to the IV computed with respect to a differnt template. (a)IV computed for Region of Interest Identifier from 1 to 27 (b)IV computed for Region of Interest Identifier from 28 to 57.

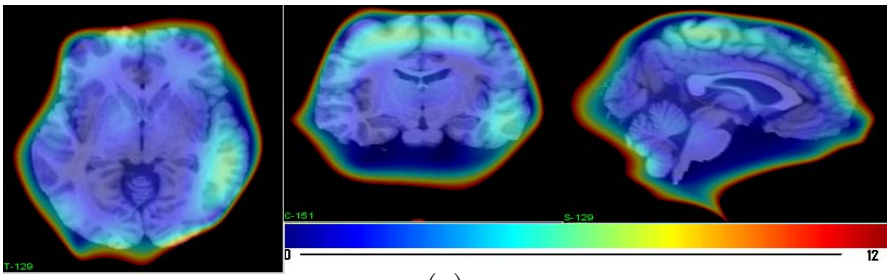
respect to rigid registration, 47 out of 57 regions have significant difference between it and Affine using IV; 55 out of 57 regions have significant difference between rigid and AIR using IV; SICLE, SLE and Demons provide significantly different, i.e., smaller IV values of all the 57 regions than rigid registration; 50 out of 57 regions have significant difference between Affine and AIR using IV; SLE, SICLE, Demons have significantly different, i.e., smaller IV over all the 57 regions than Affine; for AIR, 30 out of 57 regions have significant difference between it and SICLE, and 25 out 57 between it and SICLE; the IV for just 2 out of 57 regions have significant difference between SICLE and SLE; and Demons provides 56 out of 57 significantly different, i.e., smaller IV than SICLE and SLE.

3.5 Evaluation Results for ICE

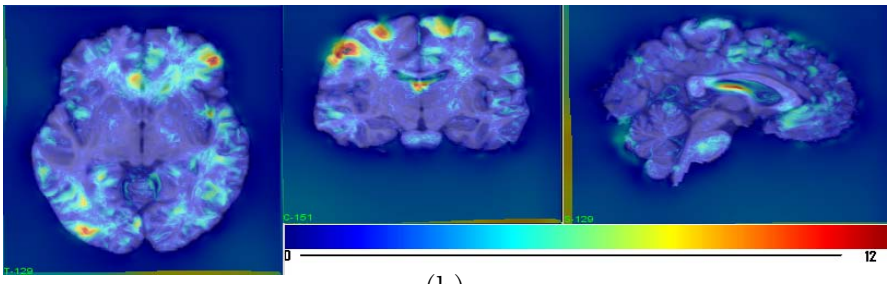
3.5.1 NA0 database

Figure 3.8 shows the inverse consistency error (ICE) for five non-rigid registration algorithms. The ICE is shown for the transformation from data set na01 to na03 concatenated with the transformation from na03 to na01. In (a)-(e) of Figure 3.8, the locations of ICE are shown by superimposing the ICE image on top of the target T1 image na01. The error images are visualized using a pseudo color scale. These images are important to visualize the errors besides computing summary statistics. The localization of the errors can be analyzed and used to develop new improved algorithms.

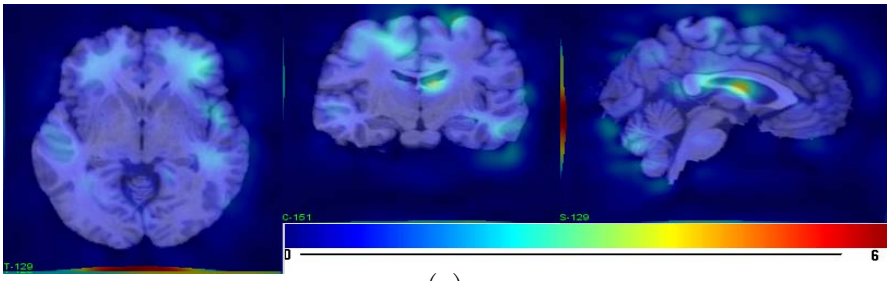
The graph in Figure 3.9 plots ICE for all 5 non-rigid registration algorithms. ICE were computed for each of the 32 ROIs, and generated from all 240 pair-wise



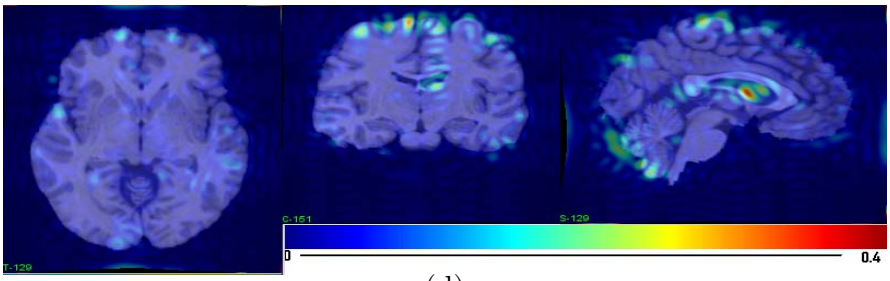
(a)



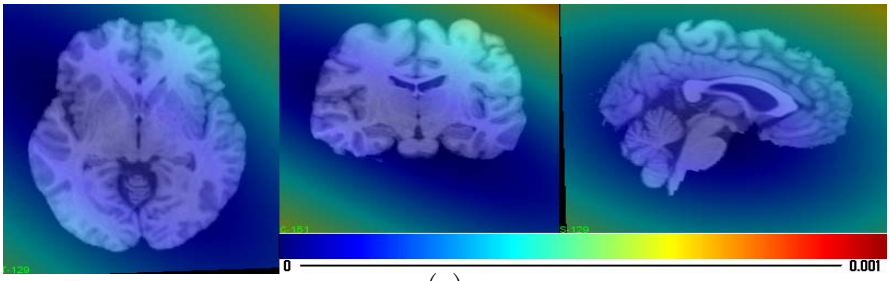
(b)



(c)



(d)



(e)

Figure 3.8: An example showing Inverse Consistency Error(color-coded map on top) superimposed on MRI image of NA0. The Target MRI image here is na01. (a)ICE for AIR (b)ICE for Demons (c)ICE for SLE (d)ICE for SICLE (e)ICE for Affine.The images are ordered from worst performance (top) to best performance (bottom).

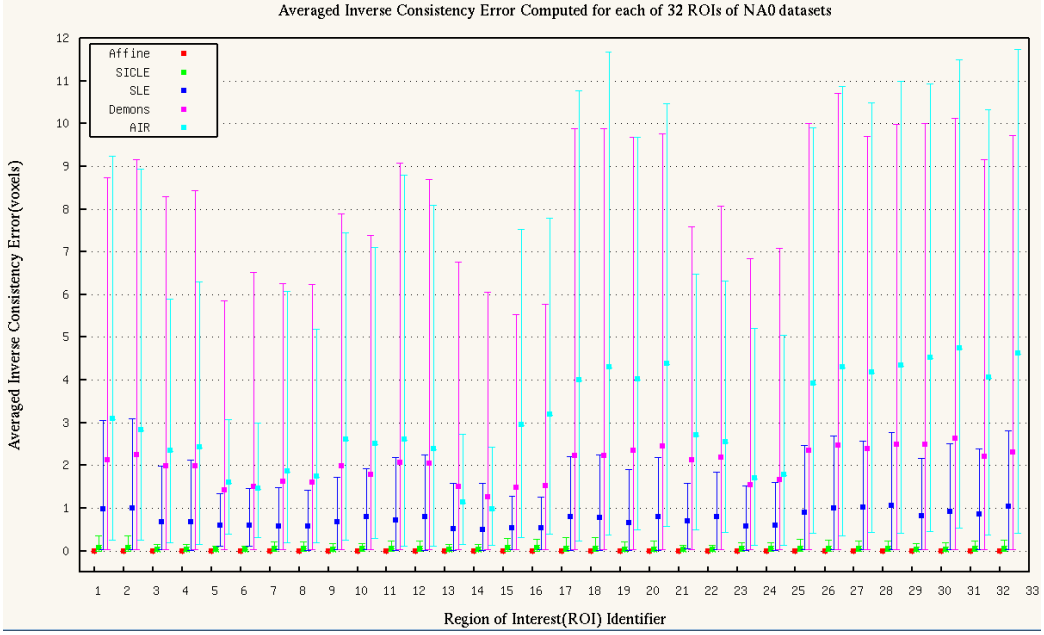


Figure 3.9: Graphs of average Inverse Consistency Error(ICE) of 16 NA0 datasets, computed for 240 transformations from five registration algorithms. Each measurement corresponds to the ICE computed with respect to a different template.

transformations of 16 T1 images in the NA0 database. In each ROI, minimum, maximum and mean voxel-wise ICE values were selected and averaged over all measurements. The value of minimum, maximum and mean ICE values in each ROI were then averaged over all ROIs, and the values are shown in table 3.6. These present the different performance of registration algorithms.

By applying Student's t-test as described, we found under the significant level 0.05: with respect to Affine, all other 4 non-rigid registration algorithms provide significantly different, i.e., bigger ICE values of all the 32 regions than it; SLE, AIR, and Demons have significantly different, i.e., bigger ICE over all the 32 regions than SICLE; for SLE, all of 32 regions have significant difference between it and Demons,

Table 3.6: Inverse Consistency Error of NA0

Registration Algorithms	Averaged min ICE in 32 ROIs	Averaged max ICE in 32 ROIs	Averaged average ICE in 32 ROIs
Affine	0.000069	0.00021	0.00014
AIR	0.3114	7.8798	2.9987
Demons	0.0321	8.2797	2.0102
SLE	0.0321	2.04	0.752
SICLE	0.0010	0.217	0.0496

as well as AIR; the ICE for 29 out of 32 have significant difference between Demons and AIR.

3.5.2 NA1 database

Figure 3.10 shows the inverse consistency error (ICE) for five non-rigid registration algorithms. The ICE is shown for the transformation from data set 3065 to 3362 concatenated with the transformation from 3362 to 3065. In (a)-(e) of Figure 3.10, the locations of ICE are shown by superimposing the ICE image on top of the target T1 image 3065. The error images are visualized using a pseudo color scale. These images are important to visualize the errors besides computing summary statistics. The localization of the errors can be analyzed and used to develop new improved algorithms.

The graph in Figure 3.11 plots the ICE for all the 5 non-rigid registration algorithms. ICE were computed for each of the 57 ROIs, and generated from all 30 pair-wise transformations of 6 T1 images in the NA1 database. In each ROI, minimum, maximum and mean voxel-wise ICE values were selected and averaged over all measurements. The value of minimum, maximum and mean ICE values in

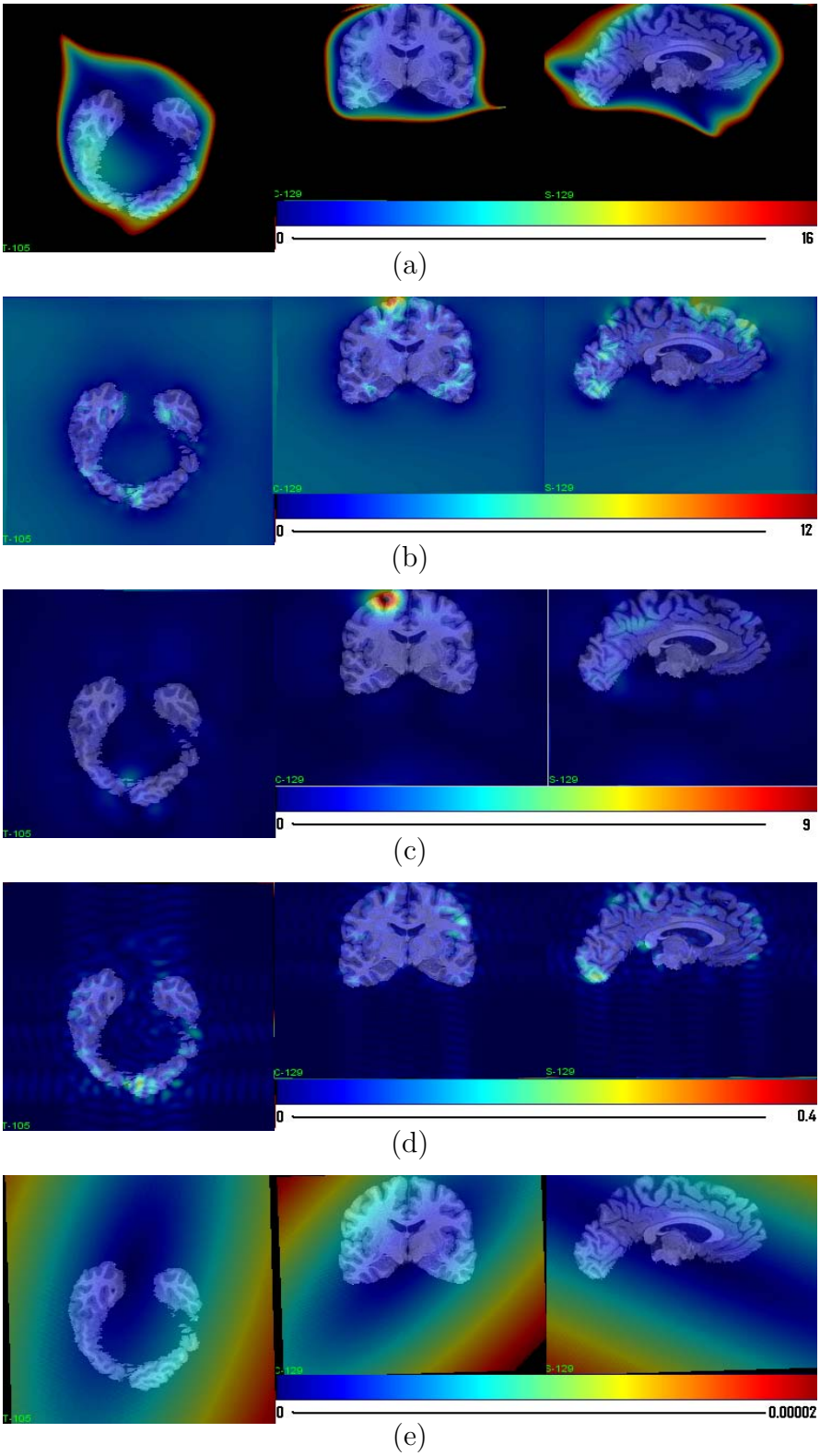


Figure 3.10: An example showing Inverse Consistency Error(color-coded map on top) superimposed on MRI image of NA1. The Target MRI image here is 3065. (a)ICE for AIR (b)ICE for Demons (c)ICE for SLE (d)ICE for SICLE (e)ICE for Affine.The images are ordered from worst performance (top) to best performance (bottom).

Table 3.7: Inverse Consistency Error 0f NA1

Registration Algorithms	Averaged min ICE in 57 ROIs	Averaged max ICE in 57 ROIs	Averaged average ICE in 57 ROIs
Affine	0.000294	0.00189	0.000624
AIR	0.3694	4.8127	1.8678
Demons	0.0580	4.8000	1.4566
SLE	0.0389	1.4124	0.4738
SICLE	0.0017	0.1487	0.0383

each ROIs were then averaged over all ROIs, and values are shown in table 3.7.

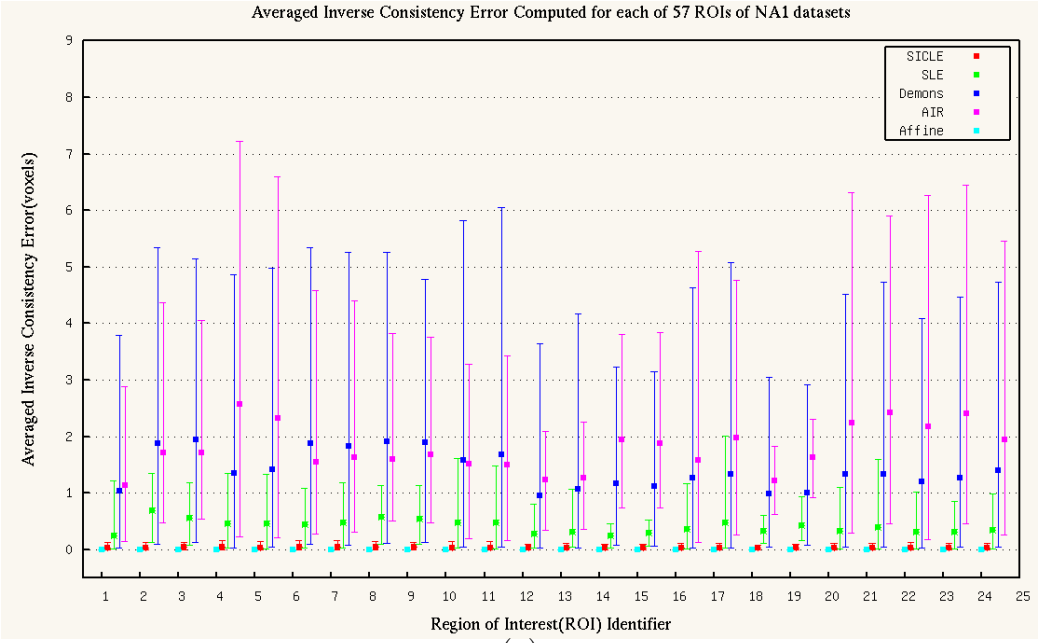
By applying Student's t-test, we found under the significant level 0.05: with respect to Affine, all other 4 registration algorithms provide significantly different, i.e., bigger ICE values of all the 57 regions than it; SLE, AIR, and Demons have significant different, i.e., bigger ICE over all the 57 regions than SICLE; for SLE, all of 57 regions have significant difference between it and Demons, as well as AIR; the ICE for 37 out of 57 regions have significant difference between Demons and AIR.

3.6 Evaluation Results for TE

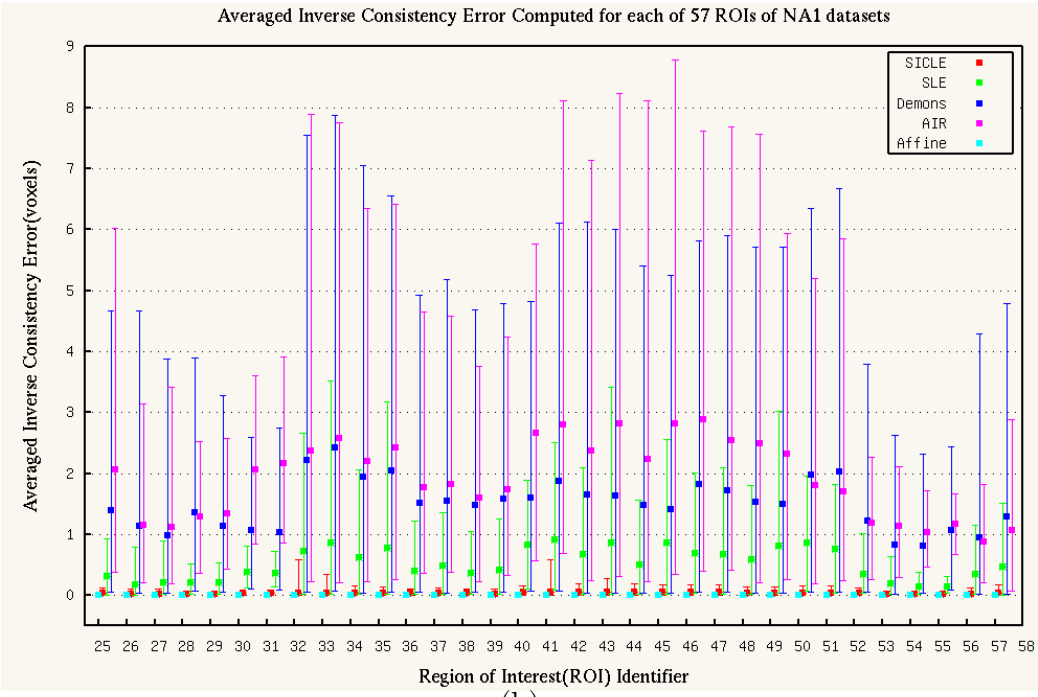
3.6.1 NA0 database

Figure 3.12 shows the transitivity error (TE) for five non-rigid registration algorithms. TE is shown for the transformation from na09 to na04 concatenated with the transformation from na04 to na01 concatenated with the transformation from na01 to na09. In (a)-(e) of Figure 3.14, the locations of TE are shown by superimposing the TE error image on top of the target T1 image 3368. The error images are visualized using a pseudo color scale.

The graph in Figure 3.13 plots TE for all the 5 non-rigid registration algo-



(a)



(b)

Figure 3.11: Graphs of average Inverse Consistency Error(ICE) of 6 NA1 datasets, computed for 30 transformations from five registration algorithms. Each measurement corresponds to the ICE computed with respect to a different template. (a)ICE computed for Region of Interest Identifier from 1 to 24 (b)ICE computed for Region of Interest Identifier from 25 to 57.

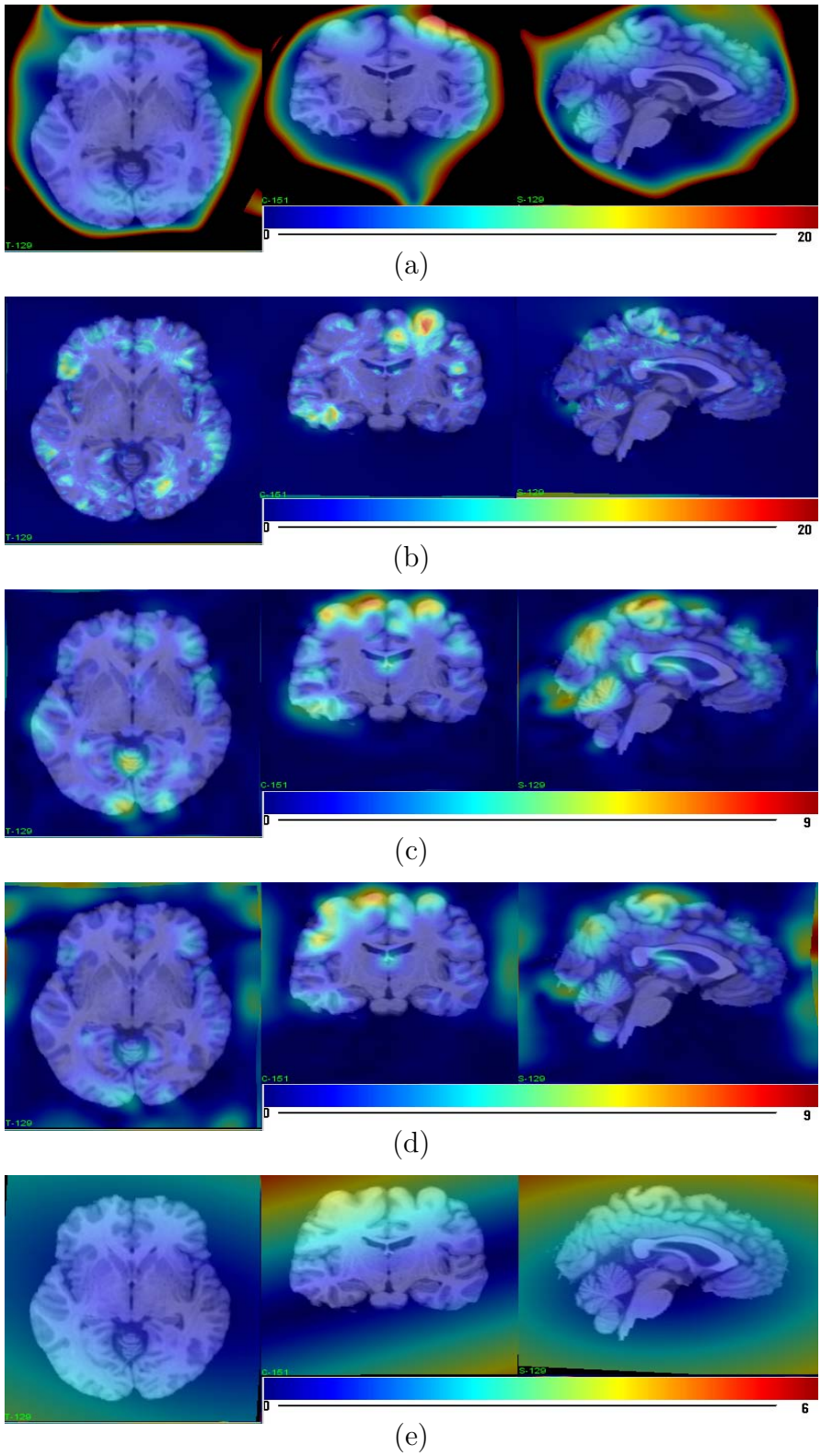


Figure 3.12: An example showing Inverse Consistency Error(color-coded map on top) superimposed on MRI image of NA0. The Target MRI image here is na01. (a)ICE for AIR (b)ICE for Demons (c)ICE for SLE (d)ICE for SICLE (e)ICE for Affine.The images are ordered from worst performance (top) to best performance (bottom).

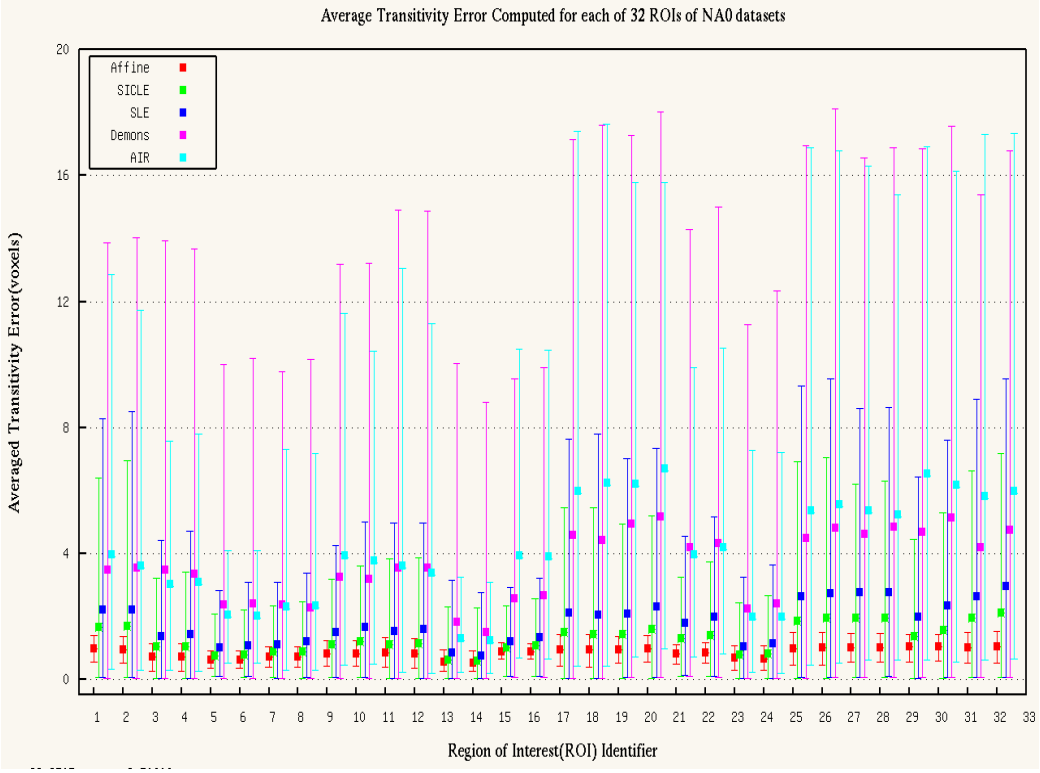


Figure 3.13: Graphs of average Transitivity Error(TE) of 16 NA0 datasets, computed for 32 ROIs and 160 transformations from five registration algorithms. This measurement correspond to TE computed with respect to different templates.

rithms. TE values were computed for each of the 32 ROIs, 16 data sets, and randomly selected 160 transformations. In each ROI, minimum, maximum and mean TE values were selected and averaged over all measurements. The values of minimum, maximum and mean TE in each ROI were then averaged over all ROIs, and the values are shown in Table 3.8.

By applying Student's t-test, we found under the significant level 0.05: with respect to Affine, 24 out of 32 regions have significant difference between it and SICLE using TE; all other 3 registration algorithms provide significantly different,

Table 3.8: Transitivity Error of NA0

Registration Algorithms	Averaged min TC in 32 ROIs	Averaged max TC in 32 ROIs	Averaged average TC in 32 ROIs
Affine	0.4436	1.2480	0.8457
AIR	0.4695	11.5922	4.0823
Demons	0.0478	14.0076	3.5942
SLE	0.0591	5.7668	1.7900
SICLE	0.0496	4.2496	1.2972

i.e., bigger TE values of all the 32 regions than it; SLE, AIR, and Demons have significant difference over all the 32 regions than SICLE using TE; for SLE, all of 32 regions have significant difference between it and Demons, as well as AIR; the ICE for 21 out of 32 have significant difference between Demons and AIR.

3.6.2 NA1 database

Figure 3.14 shows the transitivity error (TE) for five non-rigid registration algorithms. TE is shown for the transformation from 3065 to 3362 concatenated with the transformation from 3362 to 3368 concatenated with the transformation from 3368 to 3065. In (a)-(e) of Figure 3.14, the locations of TE are shown by superimposing the TE image on top of the target T1 image 3368. The error images are visualized using a pseudo color scale.

The graph in Figure 3.15 plots TE for all the 5 registration algorithms. TE were computed for each of the 57 ROIs, and generated from all 30 pair-wise transformations among those 6 data sets of NA1 database. In each ROI, minimum, maximum and mean voxel-wise TE values were selected and averaged over all measurements.

By applying Student's t-test, we found under the significant level 0.05: with

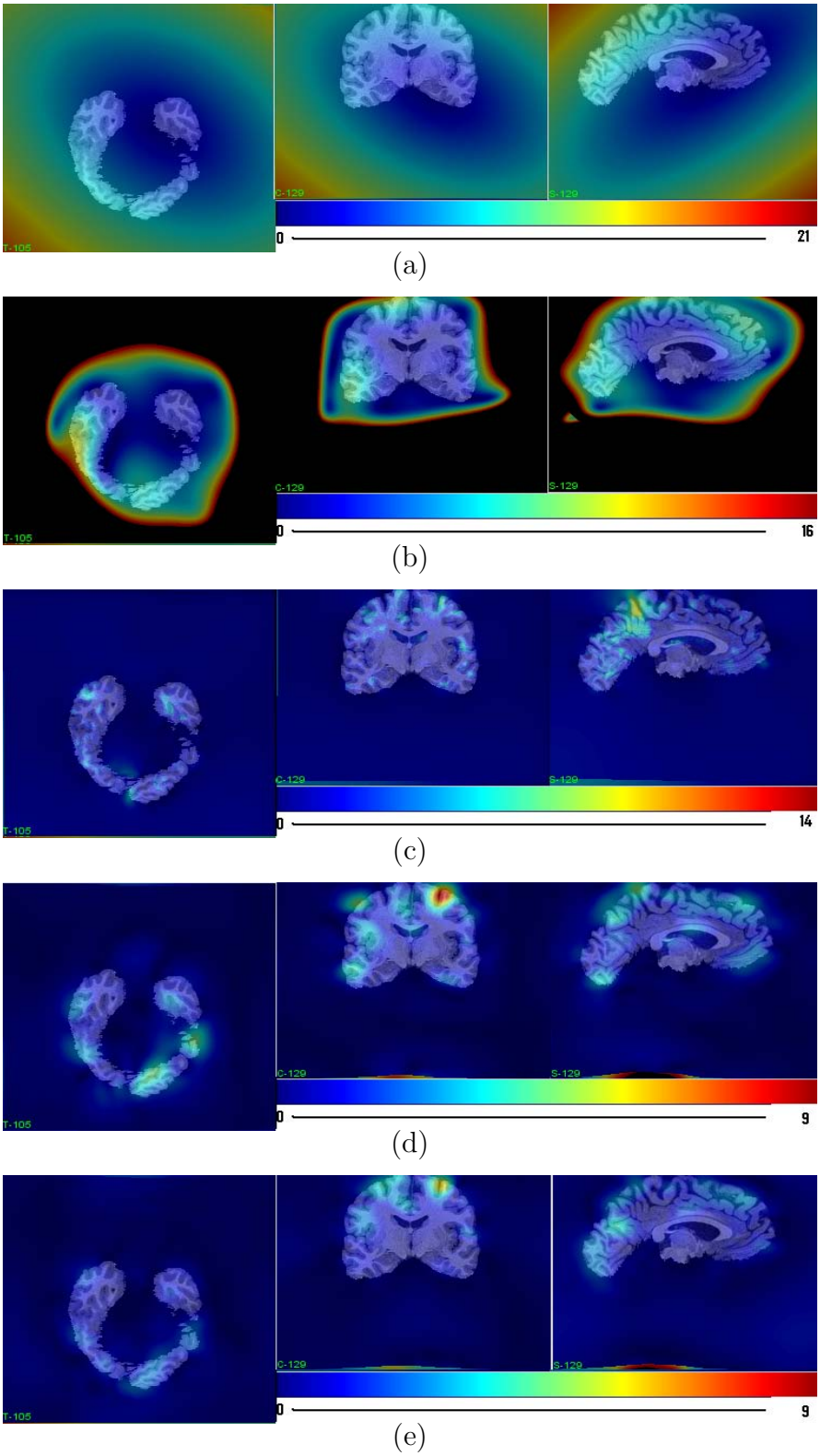
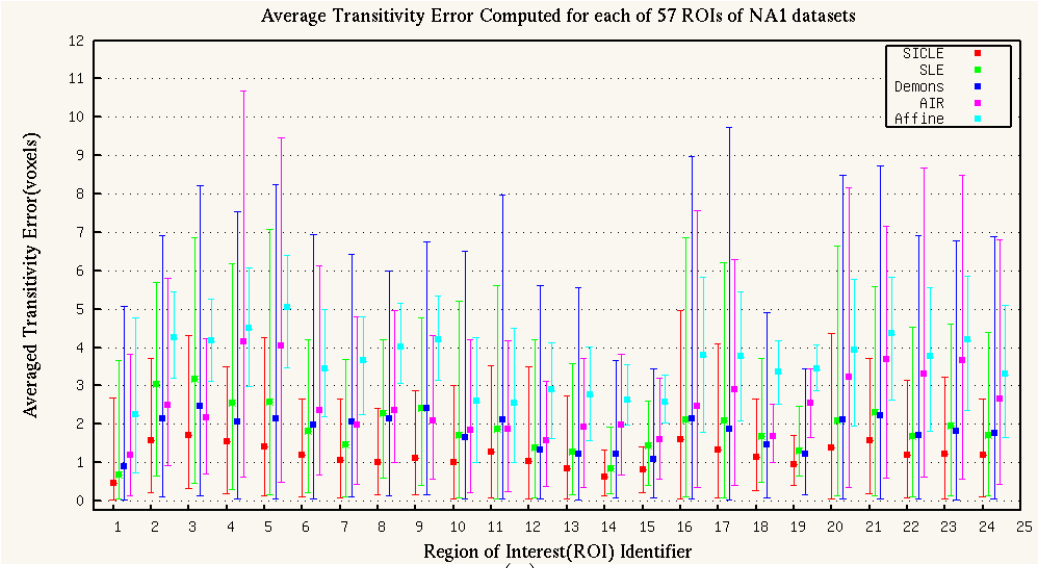
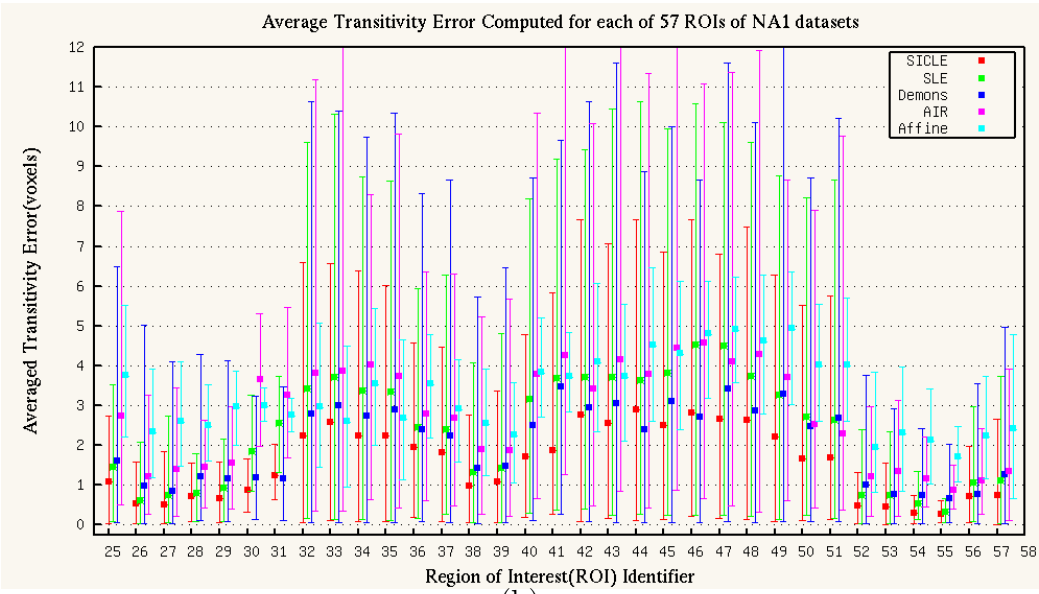


Figure 3.14: An example showing Transitivity Error(color-coded map on top) superimposed on MRI image of NA1. The Target MRI image here is 3065. (a)TC for Affine (b)TC for AIR (c)TC for Demons (d)TC for SLE (e)TC for SICLE. The images are ordered from worst performance (top) to best performance (bottom).



(a)



(b)

Figure 3.15: Graphs of average Transitivity Error(TE) of 6 NA1 datasets, computed for 20 transformations from five registration algorithms. This measurement correspond to TE computed with respect to different templates. (a)TE computed for Region of Interest Identifier from 1 to 24 (b)TE computed for Region of Interest Identifier from 25 to 57.

Table 3.9: Transitivity Error of NA1

Registration Algorithms	Averaged min TC in 57 ROIs	Averaged max TC in 57 ROIs	Averaged average TC in 57 ROIs
Affine	2.0429	4.8407	3.3822
AIR	0.5584	6.6383	2.6680
Demons	0.0698	7.0379	1.9376
SLE	0.2252	5.6020	2.1616
SICLE	0.1254	3.8082	1.3979

respect to SICLE, all other 4 registration algorithms provide significantly different, i.e., bigger TE values of all the 57 regions than it; for SLE, 33 out of 57 regions have significant difference between it and Demons; SLE provides significant difference for 42 out of 57 regions than AIR using TE, and 40 out of 57 regions than Affine using TE; More, the ICE for 41 out of 57 have significant difference between Demons and AIR; and Demons provides significant difference for 39 out of 57 regions than Affine using TE; 35 out of 57 regions have significant different TE between Affine and AIR.

3.7 Evaluation through fMRI

3.7.1 Variability of functional response under eccentricity stimulation

Figure 3.16 shows screen shots for the functional response of 6 data sets of NA1. These functional images show us the phase of eccentricity response in occipital region. This figure gives us a general overview of the variability of structural-functional corresponding relationship among subjects.

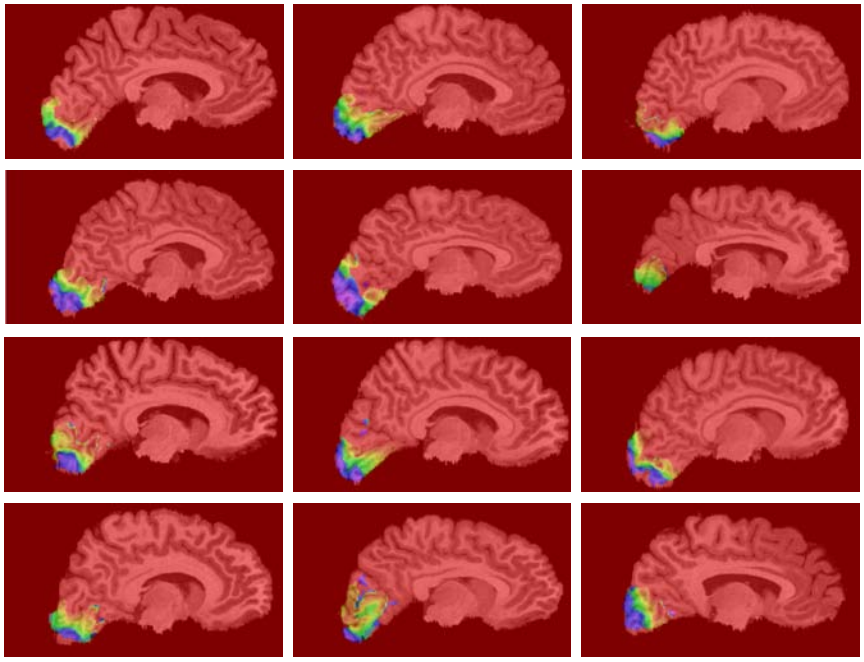


Figure 3.16: Variance of functional response among different subjects(a) phase of eccentricity in left occipital region; (b)phase of eccentricity in right occipital region.

3.7.2 Result of Method 1

The result of Averaged Normalized Cross Correlation (ANCC) is shown in Table 3.10. Here ANCC of original fMRI data sets, i.e. rigid registration, was also calculated for comparison.

Note: what also shown in Table 3.10 is relative overlap of two functional response area:

$$RO = \frac{F_i \cap F_j}{F_i \cup F_j} \tag{3.1}$$

where $F_i \cap F_j$ defines intersection area of two functional image F_i and F_j, and $F_i \cup F_j$ define is the union area of this two functional image.

And if F_j is deformed to target F_i, RO of F_i and deformed functional imageF_{ji}

Table 3.10: Averaged NCC in whole occipital region

Registration Method	RO_{F_i}	RO	NCC
Original(WIthout any registration)	47%	30%	0.603
Affine	64%	39%	0.486
AIR	73%	48%	0.498
Demons	76%	49%	0.505
SICLE	74%	47%	0.485
SLE	78%	54%	0.504

should be:

$$RO_{=}\frac{F_i \cap F_{ji}}{F_i \cup F_{ji}} \tag{3.2}$$

If we just consider the percentage of the intersection area in target image F_i, relative overlap is:

$$RO_{F_i} = \frac{F_i \cap F_{ji}}{F_i} \tag{3.3}$$

3.7.3 Result of Method 2

We just calculated NCC between fMRIs of 3362 and 3065, as well as deformed fMRI from 3362 to 3065 by different registration algorithms. Just one pair-wise registration result was evaluated, yet problems already came up.

Remind the ROI number first:

l2: left Cuneus;

l4: left Lateral Occipital Gyrus;

l6: left Lingual Gyrus;

l8: left Calcarine (pericalcarine) Region.

Table 3.11: Relative overlap of each ROI in anatomy MRI

registration method	relative overlap of each ROI							
	l2	l4	l6	l8	r3	r5	r7	r9
Original	0.142	0.164	0.146	0.121	0.096	0.178	0.189	0.140
Affine	0.146	0.204	0.201	0.155	0.092	0.185	0.208	0.140
AIR	0.175	0.253	0.248	0.183	0.170	0.238	0.264	0.203
Demons	0.259	0.344	0.333	0.250	0.228	0.314	0.367	0.278
SICLE	0.197	0.277	0.281	0.201	0.153	0.250	0.299	0.213
SLE	0.211	0.292	0.303	0.220	0.178	0.259	0.306	0.235

r3: right Cuneus;

r5: right Lateral Occipital Gyrus;

r7: right Lingual Gyrus;

r9: right Calcarine (pericalcarine) Region.

3.7.3.1 Relative overlap of each ROI

First the result of relative overlap of 8 ROIs respectively in anatomy MRI is shown in Table 3.11. This shows how two corresponding ROIs are aligned to each other by different algorithms. We also knew how large the intersection area of two corresponding ROIs from two fMRIs is, as we calculated statistics only with those functional responses in the intersection area.

3.7.3.2 Functional response

Then let us see how much response of fMRI falls in graymatter (ROI just include graymatter) of each subject, respectively, which shown in Table 3.12.

Note: Image size is $256 * 256 * 208$.

Table 3.12: Functional response

subjects	Response in whole occipital region (number of voxels)	response in graymatter (number of voxels/percentage)
3065	40008	23973/59.9%
3362	49858	27430/55.0%

Table 3.13: Intersection response area

method registration	union of response in all ROIs (number of voxels)	intersection of response in all ROIs (number of voxels)	intersection of response in each ROI (number of voxels)							
			l2	l4	l6	l8	r3	r5	r7	r9
Original	10658	5317	42	1638	636	215	72	1202	982	530
Affine	11334	5437	17	1693	746	134	59	1148	1176	464
AIR	15113	7333	10	2310	787	205	152	1910	1262	697
Demons	17684	8723	21	2794	1115	300	130	2103	1525	735
SICLE	15375	7751	25	2546	990	382	155	1782	1231	640
SLE	15718	7869	22	2739	944	418	152	1709	1218	667

3.7.3.3 Intersection of functional response

in the intersection area

Third, the values of how much response in intersection area falls in these 8 ROIs together and separately, are shown in Table 3.13. Note that many numbers are too small here, and image size is $256 * 256 * 208$.

Table 3.14: NCC in each ROI

registration	NCC in all 8 ROIs	NCC in each ROI							
		l2	l4	l6	l8	r3	r5	r7	r9
Original	0.686	-0.693	0.635	0.444	0.654	0.148	0.702	0.678	0.113
Affine	0.501	-0.098	0.310	0.351	0.541	0.018	0.386	0.552	-0.220
AIR	0.584	0.160	0.309	0.303	0.670	-0.087	0.250	0.719	0.601
Demons	0.571	-0.575	0.22	0.348	0.740	0.425	0.442	0.663	0.574
SICLE	0.574	0.322	0.297	0.386	0.750	0.068	0.285	0.700	0.582
SLE	0.602	0.449	0.310	0.304	0.786	0.151	0.303	0.706	0.541

3.7.3.4 NCC in each ROI

Finally, let us see the NCC values in 8 ROIs together and in each ROI respectively, which shown in Table 3.14.

3.8 STARD Documentation

3.8.1 STARD Documentation of NA0

The following is a checklist of the 25 items that were reported when documenting evaluation results of registration algorithms for the NIREP NA0 database. The list describes each checklist item and then gives the required documentation for NA0 database of 16 MRI images with 32 ROIs and the evaluation statistic performance of five non-rigid registration algorithms. Investigators that evaluate new registration algorithms with the NA0 database can refer back to this STARD documentation and only have to complete the items that are different. Note that not all 25 items of the STARD documentation are applicable for reporting NIREP registration results.

Item 1. Describe the purpose of this work.

This is a study of the diagnostic accuracy of non-rigid image registration algorithms.

Item 2. Describe the study aim.

The goal of this study is to establish, maintain, and endorse a standardized set of relevant benchmarks and metrics for performance evaluation of nonrigid image registration algorithms.

Item 3. Describe how the database images were selected for the evaluation population including the inclusion and exclusion criteria.

Eight male and eight female subjects were drawn at random from 240 normal volunteers who participated in functional imaging experiments of language or emotion conducted with positron emission tomography (PET) at the University of Iowa Department of Neurology. Afterward, three of these subjects were removed, at random, and replaced by three additional subjects of the same sex from ethnic minority groups who were underrepresented in the original sample. The inclusion/exclusion criteria for the functional imaging experiments were: 1) age $>= 18$*; 2) right-handed. Only subjects with scores between* $+85$ *and* $+100$*, inclusive, were included; 3) native English speaker; 4) Average intelligence or higher (estimated verbal IQ* $>= 90$ *according to the National Adult Reading Test-Revised (NART-R)); 5) free of neurological and psychiatric disease, and of development disabilities, based on interviews by a neurologist and neuropsychologist to determine neurological and psychiatric history; 6) Other medical*

*exclusions include severe hypertension or coronary artery disease, anemia, renal disease, thyroid dysfunction, and medications which affect cerebral blood flow; 6) Normal visual acuity (*20/20 *or corrected to* 20/20*); 7) No factor which contraindicates MR scanning, including pacemaker, pacemaker wires, aneurysm clip, or any electronic implant, inner ear surgery, weight over* $136kg(300lb)$*, metal embedded in soft tissue or in the eye, prosthetic eye, pregnancy (as noted in #*6 *above), or claustrophobia; 9) No participation in research within the past year involving exposure to ionizing radiation; 10) No exposure in his or her daily activity, to ionizing radiation (e.g. CT or PET technicians, etc.).*

Item 4. Describe how the participants imaged in the evaluation population were recruited.

Subjects were recruited from the Iowa City community at large, the University of Iowa Hospitals setting, and the University of Iowa using newspaper and other local advertising.

Item 5. Describe the participant sampling, i.e., was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4 or not.

The study population was not a consecutive series, but a sample drawn from a consecutive series of 240 normal volunteers who were recruited for functional imaging studies mentioned above.

Item 6. Describe the data collection procedure.

This is a retrospective study. The MRI data was collected for a previous project. See the papers [1, 2] for the details.

Item 7. Describe each evaluation statistics in detail and motivate in what context each makes sense.

See Section 2.4.1, Section 2.4.2, Section 2.4.4, Section 2.4.5, and Section 2.4.6.

Item 8. Describe the technical specifications of the registration algorithm including all parameters used to generate the results. The executable program used to generated the results will be made available if possible.

See Section 2.3.

Item 9. Describe the definition of and rationale for the units, cutoffs and categories used to report the results will be fully described.

Not Applicable. There is no reference standard. Parameters and cutoffs are chosen to maximize the performance of the registration algorithm. Alternatively, the registration results produced using different parameters and cutoffs can be compared with the NIREP framework.

Item 10. Describe the number, training and expertise of the personal segmenting and approving segmentations and labellings for the evaluation database.

Tissue segmentation was performed by Joel Bruss BA, research assistant II, under the supervision of neurologist Thomas J. Grabowski, MD. Mr. Bruss has more than five years of experience with anatomical analysis and structural image

processing. Dr. Grabowski is Associate Professor of Neurology and Radiology, and has extensive expertise in neuroanatomy, neuroimaging, and image processing. Cortical parcellation was performed by Joel Bruss BA and John Allen PhD, under the supervision of neurologists Hanna Damasio MD and Thomas J. Grabowski, MD. Combination of tissue segmentation and cortical parcellation was performed by predoctoral studient Xiujuan Geng, and research assistant II Joel Bruss BA, under the supervision of Thomas J. Grabowski, MD.

Item 11. Describe whether or not the readers of the index tests and reference standard blinded to the results of the other tests.

All test statistics are performed by the NIREP software and applied to each registration method the same way. Therefore the tests are blinded to the outcomes of other tests.

Item 12. Describe the statistical methods used to quantify the uncertainty, i.e., the criteria used to denote statistically significant differences between registration algorithms will be described.

The statistical significance of the differences in the evaluation metrics were assessed using Students t-test. For each metric, the t-test was applied for each region of interest. Alpha was set equal to 0.05*. A two-tailed equal variance t-test was used for relative overlap and intensity variance. A two-tailed paired t-test was used for inverse consistency error and transitivity error. A significant difference was reported if the p value was less than* 0.05*.*

Item 13. Describe the methods for calculating test reproducibility.

No reproducibility studies were done for the experiments reported in this paper.

Item 14. Report the dates that the study was done.

The SICLE registrations were started on October 27, 2005 and were completed on March 21, 2005. The SLE registrations were started on February 3, 2006 and were completed on February 12, 2006. The AIR resgistrations were started on October 30, 2007 and completed on December 20, 2007. Demons registrations were started on March, 2008 and completed on May, 2008. Affine registrations were started on October, 2008 and completed November,2008.

Item 15. Report the clinical and demographic characteristics of the study population (e.g., age, sex, spectrum of presenting symptoms, comorbidity, current treatments, recruitment centers).

See Section 2.2.1.

Item 16. Describe the number of data sets that satisfied the criteria for inclusion that were or were not included in the evaluation metrics. For example, it is possible that some registrations may be excluded from analysis if an algorithm fails to produce a satisfactory registration.

No data sets were excluded from this analysis.

Item 17. Report the time interval from the index tests to the reference standard, and any treatment administered between them.

Not Applicable. The MRI scans were collected once per subject. Thus, there was no treatment administered between them.

Item 18. Report the distribution and severity of any disease (define criteria) in the evaluation population.

Only normal subjects participated in this experiment.

Item 19. Report a cross tabulation/sorted-graph of the results of the evaluation metrics with respect to each segmented/labeled region in the reference/template image volume.

See Section3.1.1,Section3.2.1, Section3.4.1, Section3.5.1 and Section3.6.1.

Item 20. Report any adverse events from performing the index tests or the reference standard.

Not applicable for the NA0 evaluation database.

Item 21. Report the estimates of diagnostic accuracy and measures of statistical uncertainty of the evaluation metrics (e.g., 95% confidence intervals).

See Section3.1.1,Section3.2.1, Section3.4.1, Section3.5.1 and Section3.6.1.

Item 22. Describe how indeterminate results, missing responses and outliers of the index tests were handled.

The evaluation metrics were applied to all the data sets and all results were used.

Item 23. Report estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.

Not applicable to the current study.

Item 24. Report estimates of test reproducibility, if done.

The evaluation metrics were fixed and give the same answer each time they are applied.

Item 25. Report the clinical applicability of the study findings.

Colocalization of cortical anatomy is an essential prerequisite to computational neuroanatomical studies, e.g. those that seek to identify structural brain correlates of conditions that are not associated with overt pathology (e.g. developmental dyslexia), or those that seek to identify structural brain correlates of genetic factors. Colocalization of cortical anatomy across normal individuals is also an important prerequisite to studies of brain function, such as functional MRI studies of cognition.

3.8.2 STARD Documentation of NA1

The following is a checklist of the 25 items that were reported when documenting evaluation results of registration algorithms for the NIREP. The list describes each checklist item and then gives the required documentation for NA1 database of 18 MRI images with 57 ROIs and the evaluation statistic performance of five non-rigid registration algorithms described in Section2.3. Investigators that evaluate new registration algorithms with the NA1 database can refer back to this STARD docu-

mentation and only have to complete the items that are different. Note that not all 25 items of the STARD documentation are applicable for reporting NIREP registration results.

Item 1. Describe the purpose of this work.

This is a study of the diagnostic accuracy of non-rigid image registration algorithms.

Item 2. Describe the study aim.

The goal of this study is to establish, maintain, and endorse a standardized set of relevant benchmarks and metrics for performance evaluation of nonrigid image registration algorithms.

Item 3. Describe how the database images were selected for the evaluation population including the inclusion and exclusion criteria.

See Section2.2.2.1.

Item 4. Describe how the participants imaged in the evaluation population were recruited.

These 18 subjects will be recruited from 1) the University Hospital setting, recruited through local advertising; 2) local communities; in Iowa, recruited through newspaper advertising (and with special targeting of minority populations); 3)a dataset of approximately 80 normal control subjects who agree to be contacted for future studies under IRB# 2000303002 - Mechanisms of Perimetric Variability - Michael Wall, PI. 4) college students at the University of

Iowa, through local advertising; 5) Referral from colleague - Patients of Michael Wall, MD. Dept. of Neurology, who is a coinvestigator on this project; 6) Other-Word of mouth.

Item 5. Describe the participant sampling, i.e., was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4 or not.

Yes, the study population was a consecutive series of participants defined by the selection creteria in item 3 and 4.

Item 6. Describe the data collection procedure.

This is a prospective study.

Item 7. Describe each evaluation statistics in detail and motivate in what context each makes sense.

See Section2.4.1, Section2.4.2, Section2.4.4, Section2.4.3, Section2.4.5, and Section2.4.6.

Item 8. Describe the technical specifications of the registration algorithm including all parameters used to generate the results. The executable program used to generated the results will be made available if possible.

See Section 2.3.

Item 9. Describe the definition of and rationale for the units, cutoffs and categories used to report the results will be fully described.

Not Applicable. There is no reference standard. Parameters and cutoffs are chosen to maximize the performance of the registration algorithm. Alternatively, the registration results produced using different parameters and cutoffs can be compared with the NIREP framework.

Item 10. Describe the number, training and expertise of the personal segmenting and approving segmentations and labellings for the evaluation database.

Tissue segmentation was performed by Joel Bruss BA, research assistant II, under the supervision of neurologist Thomas J. Grabowski, MD. Mr. Bruss has more than five years of experience with anatomical analysis and structural image processing. Dr. Grabowski is Associate Professor of Neurology and Radiology, and has extensive expertise in neuroanatomy, neuroimaging, and image processing. Cortical parcellation was performed by Joel Bruss BA under the supervision of neurologists Hanna Damasio MD and Thomas J. Grabowski, MD. Combination of tissue segmentation and cortical parcellation was performed by predoctoral studient Ying Wei, and research assistant II Joel Bruss BA, under the supervision of Thomas J. Grabowski, MD.

Item 11. Describe whether or not the readers of the index tests and reference standard blinded to the results of the other tests.

All test statistics are performed by the NIREP software and applied to each registration method the same way. Therefore the tests are blinded to the outcomes of other tests.

Item 12. Describe the statistical methods used to quantify the uncertainty, i.e., the criteria used to denote statistically significant differences between registration algorithms will be described.

The statistical significance of the differences in the evaluation metrics were assessed using Students t-test. For each metric, the t-test was applied for each region of interest. Alpha was set equal to 0.05*. A two-tailed equal variance t-test was used for relative overlap and intensity variance. A two-tailed paired t-test was used for inverse consistency error and transitivity error. A significant difference was reported if the p value was less than* 0.05*.*

Item 13. Describe the methods for calculating test reproducibility.

No reproducibility studies were done for the experiments reported in this paper.

Item 14. Report the dates that the study was done.

All the registrations were started on September, 2008 and were completed on December, 2008.

Item 15. Report the clinical and demographic characteristics of the study population (e.g., age, sex, spectrum of presenting symptoms, comorbidity, current treatments, recruitment centers).

See Section2.2.2.1.

Item 16. Describe the number of data sets that satisfied the criteria for inclusion that were or were not included in the evaluation metrics. For example, it is

possible that some registrations may be excluded from analysis if an algorithm fails to produce a satisfactory registration.

Six datasets of NA1 were included for this analysis.

They are 3065,3362,3368,3402,3407,3413.

Item 17. Report the time interval from the index tests to the reference standard, and any treatment administered between them.

Not Applicable. The MRI scans were collected once per subject. Thus, there was no treatment administered between them.

Item 18. Report the distribution and severity of any disease (define criteria) in the evaluation population.

Only normal subjects participated in this experiment.

Item 19. Report a cross tabulation/sorted-graph of the results of the evaluation metrics with respect to each segmented/labeled region in the reference/template image volume.

See Section3.1.1,Section3.2.1, Section3.3,Section3.4.1, Section3.5.1 and Section3.6.1.

Item 20. Report any adverse events from performing the index tests or the reference standard.

Not applicable for the NA1 evaluation database.

Item 21. Report the estimates of diagnostic accuracy and measures of statistical uncertainty of the evaluation metrics (e.g., 95% confidence intervals).

See Section3.1.1,Section3.2.1, Section3.3,Section3.4.1, Section3.5.1 and Section3.6.1.

Item 22. Describe how indeterminate results, missing responses and outliers of the index tests were handled.

The evaluation metrics were applied to 6 data sets and all results were used.

Item 23. Report estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.

Not applicable to the current study.

Item 24. Report estimates of test reproducibility, if done.

The evaluation metrics were fixed and give the same answer each time they are applied.

Item 25. Report the clinical applicability of the study findings.

See Chapter5.

CHAPTER 4
DISCUSSION

4.1 Comparison between registration performance with respect to ARO

We know that bigger ARO value, the better the performance of a registration algorithm. As shown in Figure 3.1, Figure 3.2, Table 3.1.1 and Table 3.1.2, the overall best ARO statistic result is given by Demons algorithm. Also SLE and SICLE perform well too. From both the ARO statistic graph and numbers, SLE performs slightly better than SICLE. AIR and Affine work worse than these former three, yet there are still some improvements than rigid registration.

ARO statistic values range from nearly 0 to 0.55 in most ROIs, except ROIs 52-57 which already have large RO after rigid registration. The reason that these numbers are low is due partly to complexity of brain cortex and partly to the small ROIs that were used and especially ROIs just include graymatter. One advantage of these low ARO numbers is that it can serve as a good bench mark for future more advanced registration algorithms.

4.2 Comparison between registration performance with respect to normalized ROI Overlap

Figure 3.3 and Figure 3.4 shows that Affine and AIR have improvements than rigid registration, as there are more complete overlap regions(red color). SLE and SICLE have improvements over these two again. Moreover, Demons performs best

as we can see most red color which means complete overlap of all the corresponding ROIs with target image in Demons' result.

4.3 Comparison between registration performance with respect to alignment of calcarine sulci

Figure 3.5 shows after non-rigid registrations all the calcarien sulci point sets of other subjects are much closer to that of target subject than rigid registration result. We can see clearly all the other colors are closer to Yellow. Demons always perform best in alignment of these calcarine sulci. SLE performs better than AIR and SICLE, though these two win some of the performance measures as well. Affine has worst performance among these 5 registration algorithms but it still have an improvement than rigid registration.

4.4 Comparison between registration performance with respect to IV

Figure 3.6, Figure 3.7, Table 3.4.1 and Table 3.4.2 show that Demons outperforms the other 4 non-rigid registration algorithms with respect to IV statistic. SLE algorithm achieves a slightly smaller average intensity variance than SICLE. AIR performs much better than Affine in IV statistic. And Affine registration algorithm is sometimes even bad than rigid registration. This is because Affine just apply a simple rotation, translation or scaling here and this might cause worse intensity variance.

4.5 Comparison between registration performance with respect to ICE

Figure 3.8, Figure 3.9, Figure 3.10, Figure 3.11, Table 3.6 and Table 3.7 show that the SICLE algorithm has significantly less ICE than the SLE, Demons and

AIR. ICE statistic between SLE and SICLE shows that even though there is a slight drawback for adding the inverse consistency constraint like the results shown in RO and IV statistics, it is more than made up for by reducing ICE. The ICE of Affine is much significantly smaller than the other registration algorithms, even SICLE. This should result from the fact that Affine registration done by AIR5.2.5 gives two forward and reverse transformation matrixes as the inverse of each other. So that the ICE result of Affine is almost zero.

4.6 Comparison between registration performance with respect to TE

Likewise, the SICLE algorithm has less TE than the SLE, Demons and AIR. Figure 3.12, Figure 3.13, Figure 3.14 and Figure 3.15 and Table 3.8 and Table 3.9show that the inverse consistency constraint also reduced the transitivity error produced by the SICLE algorithm compared to the other registration algorithms. With respect to TE, evaluation results of NA1 and NA0 give different evaluation results on Affine. One is the smallest and the other is the biggest among all registration algorithms. This means small ICE does not guarantee small TE. Yet if TE is small, ICE should be small in the other hand. Take NA1 database for instance. Though Affine performs perfect in ICE statistic, yet as it is a very simple registration algorithm and has very less consideration of transitivity property of registration transformations, so it gives the worst TE performance among all these five registration algorithms.

4.7 Analyze of the evaluation results of different registration algorithms

From the evaluation results of 6 statistics shown in previous sections, we can make some conclusions. First we can see affine and AIR perform worse in almost all the aspects than SICLE, SLE and demons, as they have very low dimensional transformation. The transformation of affine and AIR has 12 parameters and 168 parameters respectively, whereas transformation of demons is presented by vector field image and has very high dimensional parameterization (about $4*10^7$ in the $256*256*300$ image). Transformations of SICLE and SLE are parameterized by Fourier series, which also have much higher parameterization than AIR and affine. The higher dimensional transformations can better model correspondences between anatomies with complex differences than lower dimensional transformations. Moreover, demons implemented in this work performs best in all similarity-based statistics, yet nearly the worst for ICE and TE, as it has less constraint that SICLE and SLE on transformation properties. This demonstrate that there is a trade-off between adding more constraint for improving transformation properties and getting smaller similarity-based error. The fact that one registration algorithm produced the best result for one criterion and nearly the worse for another, illustrates the need to use multiple evaluation criteria. Also, Then SICLE has nearly best performance on ICE and TE statistic evaluation. This shows us the inverse consistency constraint reduced the TE as well. SLE and SICLE algorithms are nearly identical except that the SICLE algorithm has an additional inverse consistency constraint. From this fact we demonstrate again that the trade-off of adding the extra inverse consistency constraint is that the registration

is a little bit worse with respect to the similarity-based statistics. Finally, if users would like to receive recommendation for choice of algorithm for their specific use. We can suggest users to choose demons among these five registration algorithms if similarity-based statistics are more concerned in their application and choose SICLE if good transformation-based properties are more important.

4.8 Discussion of registration of evaluation through functional image

Results of the evaluation by fMRI do not show a good trend of NCC among different registration algorithms. Because fMRI included in this work is the phase value of eccentricity, the phase value is supposed to change along calcarine sulcus and follow this trend: from the starting point of calcarine sulcus to the joint point of calcarine sulcus and parietooccipital fissure, functional response corresponds from the center to the outside of the visual field in eccentricity stimulation. Yet the result shows that there should be a variability of this trend among different individuals.

To find more details of the variability which might cause the absence of a trend, we picked up the phase values along the calcarine sulcus to show that the phase value trend along calcarine sulcus varies among different individuals. Here we chose the starting point of calcarine sulcus in those slices (in coronal plane) which have calcarine sulcus point sets outlined, instead of the bottom of calcarine sulcus. The reasons are: (1) the bottom of calcarine sulcus always has braches and is more complicated than the starting point; (2) while moving towards the end of calcarine sulcus, sometimes the starting point still has a response whereas the bottom point does not have a response already.

Figure 4.1 shows an example of the starting point we chose of calcarine sulcus. The red line is the calcarine sulcus in coronal plane. And green point is the starting point we chose to represent the calcarine sulcus in each slice. The numbers show there are six continuous slices, and the situation here is a little bit complicated. There are two branches. In Slice 1 we can see a small new branch comes up in additional to our original one. We still chose the starting point of the original one and ignore the new small one till these two branches joint to be one.

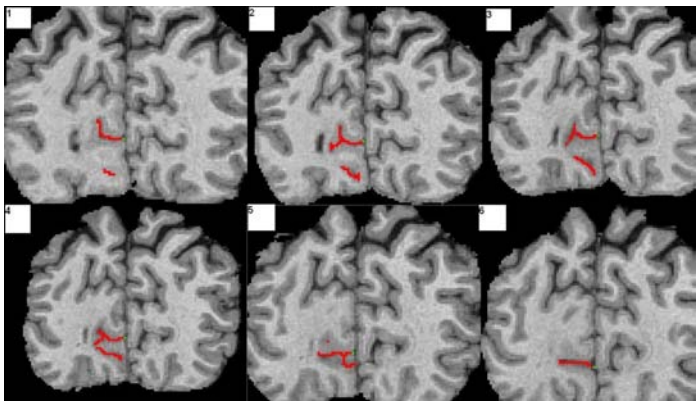


Figure 4.1: The starting point of Calcarine sulcus chosen in each slice

Next we calculated the phase value of each point by taking the averaged phase value in its neighborhood with a radius 1. Each point represented the phase in one position of calcarine sulcus, and by combining them, we saw how the phase value of eccentricity change along calcarine sulcus. All the lengths of calcarine sulcus are normalized to a same length.

Figure 4.2 shows the trend of phase value in eccentricity response along cal-

carine sulcus in left brain among different subjects. It is kind of follow the trend that along calcarine sulcus: the phase values are decreased, i.e., from center to the outsider in visual field under eccentricity stimulation (+pi represents center and -pi represents most-outsider). Yet some subjects have a sudden sharp increase in phase value (as 3065, red line).

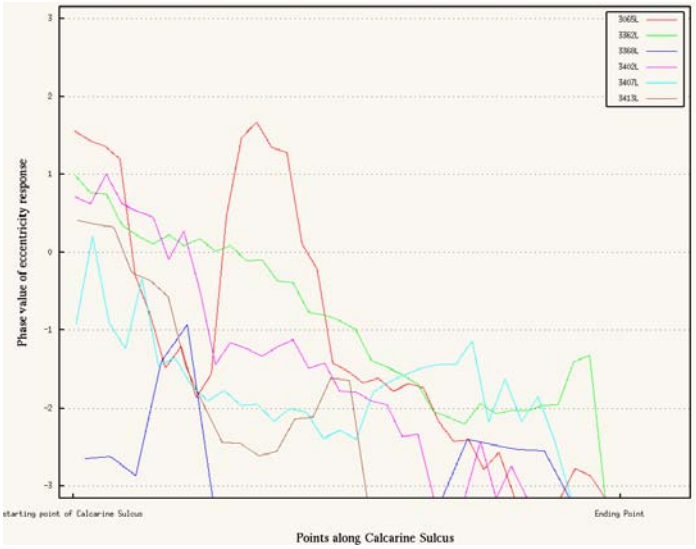


Figure 4.2: Trend of phase value in eccentricity response change along Calcarine Sulcus in left brain

Figure 4.3 shows the trend of phase value in eccentricity response along calcarine sulcus in right brain among different subjects. This time is very complicated and there seems no apparent trend here. More, in some subjects, the phase value increase first and decrease, which should not be correct.

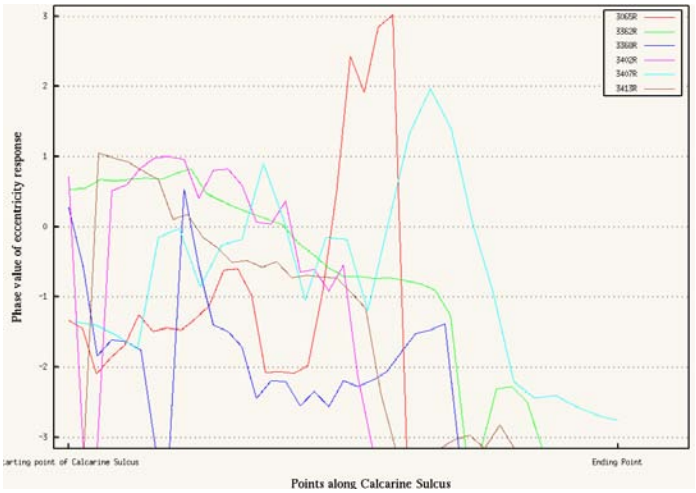


Figure 4.3: Trend of phase value in eccentricity response change along Calcarine Sulcus in right brain

CHAPTER 5
CONCLUSION

Image registration is important for many applications. Non-rigid image registration is a more general approach than the widely used affine and rigid methods, but requires more complex methodology and computational effort to implement. We have started the NIREP to develop software tools and provide shared image evaluation databases for rigorous testing of non-rigid image registration algorithms.

In this thesis work, NIREP NA0 and NA1 databases have been established and released to the research community through XNAT central. To show the ability to evaluate the registration algorithms by these two common databases, the performance of five non-rigid registration algorithms (affine, AIR, demons, SLE and SICLE) were evaluated using 22 images from two NIREP neuroanatomical evaluation databases. Six evaluation statistics (relative overlap, intensity variance, normalized ROI overlap, alignment of calcarine sulci, inverse consistency error and transitivity Error) were used to evaluate and compare registration performances. Tables, graphs, images and analysis reports were produced to help interpret registration performances and compare registration results. The results suggest that the demons registration algorithm produced the best registration results with respect to the relative overlap statistic but produced nearly the worst registration result with respect to the inverse consistency statistic. The fact that one registration algorithm produced the best result for one criterion and nearly the worst for another, illustrates the need to use multiple evaluation

criteria. To conclude, the evaluation results show that SICLE, SLE and demons work better in almost all the aspects than those registration algorithms which have lower dimensional transformations, such as AIR and affine. All these non-rigid registration results have improved the results of the rigid registration. This work illustrates the need to archive registration algorithms' performances so that investigators can choose the best registration algorithm based on their own needs. Archiving registration results can also help drive future innovation in the non-rigid image registration. This work presents that people can download our databases to evaluate their own registration algorithms through these common databases by NIREP.

In this thesis, we focus on the description of the NIREP evaluation framework and show how it can be used to evaluate non-rigid registration algorithms. This evaluation framework can be easily extended by adding more statistics and additional evaluation databases.

CHAPTER 6
PROBLEMS AND FUTURE WORK

For each pair of MR images, a transformation to deform the source image to the target image was computed by registration algorithms and then stored in a transformation database. With respect to transformation format, there is a tradeoff between storage and computation time. For example, AIR software will generate .air transformation file for Affine and .warp for AIR, which record 12 parameters and 168 parameters for registration models respectively. So that we need to select different transformation reader to read these transformation files. Or we can use space to save computation time, that is, to save transformation in displacement image-vector image format. Vector image is used in ITK for a lot of registration algorithms and has more generality. So that if the transformation format of a new registration algorithm does not have the corresponding reader in the NIREP software, displacement images provided by user can be used instead.

Also, we plan to expand and enrich the non-rigid registration evaluation MR brain database by including MR data sets with functional markers (fMRI). Yet the preliminary studies show that the evaluation results can give few trend among algorithms and cannot tell which algorithm performs better than others. We need to revise the statistic of fMRI to evaluate registration algorithms. In the future, we plan to find a better way to evaluate the variance of functional response among individuals.

Following the evaluation procedure presented in this thesis, we can evaluate

new registration algorithm by adding more standardized data sets, new statistics and new evaluation results of other algorithms. The new evaluation results will be available to the research community through the standardized documentation. Different evaluation results enable generalizable conclusions to be drawn to predict how well a certain algorithm may perform on a different image modality, different body organ, different experimental setup, different parameters, on animal data, or using an additional constraint.

REFERENCES

[1] John S. Allen, Hanna Damasio, and Thomas J. Grabowksi. Normal neuroanatomical variation in the human brain: An mri-volumetric study. *American Journal of Physical Anthropology*, 118:341–358, 2002.

[2] John S. Allen, Hanna Damasio, Thomas J. Grabowski, Joel Bruss, and Wei Zhang. Sexual dimorphism and asymmetries in the gray white composition of the human cerebrum. *NeuroImage*, 18:880–894, 2003.

[3] Y. Amit. A non-linear variational problem for image matching. *SIAM Journal on Scientific Computation*, 15(1):207–224, January 1994.

[4] P.M. Bossuyt, J.B. Reitsma, D.E. Bruns, C.A. Gatsonis, P.P. Glasziou, L.M. Irwig, J.G. Lijmer, D. Moher, D. Rennie, and H.C.W. de Vet. Towards complete and accurate reporting of studies of diagnostic accuracy: The stard initiative. *Clinical Chemistry*, 49(1):1–6, Jan. 2003.

[5] P.M. Bossuyt, J.B. Reitsma, D.E. Bruns, C.A. Gatsonis, P.P. Glasziou, L.M. Irwig, D. Moher, D. Rennie, H.C.W. de Vet, and J.G. Lijmer. The stard statement for reporting studies of diagnostic accuracy: Explanation and elaboration. *Clinical Chemistry*, 49(1):7–18, Jan. 2003.

[6] K.W. Bowyer, M.H. Loew, H.S. Stiehl, and M.A. Viergever. Report of the dagstuhl seminar 01111. In *Methodology of Evaluation in Medical Image Computing*, volume Report 301, Wadern, Germany, 2001. IBFI gem. GmbH, Schloss Dagstuhl, D-66687. http://www.dagstuhl.de/01111/Report/.

[7] Kunlin Cao. Local lung tissue expansion analysis based on inverse consistent image registration. Master's thesis, Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, IA 52242, May 2008.

[8] R. Castillo, E. Castillo, R. Guerra, V.E. Johnson, T. McPhail, A.K. Garg, and T Guerrero. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Phys Med Biol*, 54:1849–1870, 2009.

[9] Gary E. Christensen, Xiujuan Geng, Jon G. Kuhl, Joel Bruss, Thomas J. Grabowski, John S. Allen, Imran A. Pirwani, Michael W. Vannier, and Hanna Damasio. Introduction to the non-rigid image registration evaluation project (nirep). In *3rd International Workshop on Biomedical Image Registration*, LCNS 4057, pages 128–135. Springer-Verlag, July 2006.

[10] G.E. Christensen. Consistent linear-elastic transformations for image matching. In A. Kuba and M. Samal, editors, *Information Processing in Medical Imaging*, LCNS 1613, pages 224–237, Berlin, June 1999. Springer-Verlag.

[11] G.E. Christensen and H.J. Johnson. Consistent image registration. *IEEE Trans. Med. Imaging*, 20(7):568–582, July 2001.

[12] G.E. Christensen and H.J. Johnson. Invertibility and transitivity analysis for nonrigid image registration. *Journal of Electronic Imaging*, 12(1):106–117, Jan. 2003.

[13] G.E. Christensen, S.C. Joshi, and M.I. Miller. Volumetric transformation of brain anatomy. *IEEE Trans. on Med. Imaging*, 16(6):864–877, December 1997.

[14] G.E. Christensen, R.D. Rabbitt, and M.I. Miller. 3D brain mapping using a deformable neuroanatomy. *Physics in Medicine and Biology*, 39:609–618, 1994.

[15] H. T. Croft, K. J. Falconer, and R. K. Guy. *Unsolved Problems in Geometry*. Springer-Verlag, New York, 1991.

[16] W.R. Crum, O. Camara, D. Rueckert, K.K Bhatia, M. Jenkinson, and D.L.G. Hill. Generalized overlap measures for assessment of pairwise and groupwise image registration and segmentation. In *MICCAI 2005*, pages 99–106. Springer, 2005.

[17] W.R. Crum, T. Hartkens, and D.L.G. Hill. Non-rigid imgag registration: theory and practice. *The British Journal of Radiology*, 77:140–153, 2004.

[18] R.S. Desikan, F. Segonne, B. Fischl, B.T. Quinn, B.C. Dickerson, D. Blacker, R.L. Buckner, A.M. Dale, R.P. Maguire, B.T. Hyman, M.S. Albert, and R.J. Killiany. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *NeuroImage*, 31:968–80, 2006.

[19] Xiujuan Geng, Gary E. Christensen, Hong Gua, Thomas J. Rossa, and Yihong Yanga. Implicit reference-based group-wise image registration and its application to structural and functional mri. *NeuroImage*, 2009.

[20] Guido Gerig, Matthieu Jomier, and Miranda Chakos. Valmet: A new validation tool for assessing and improving 3d object segmentation. In Wiro J. Niessen and Max A. Viergever, editors, *MICCAI 2001*, volume LNCS 2208, pages 516–528. Springer, 2001.

[21] Tristan Glatard, Xavier Pennec, and Johan Montagnat. Performance evaluation of grid-enabled registration algorithms using bronze-standards. In *MICCAI 2006*, pages 152–160. Springer, 2006.

[22] P. Hellier, C. Barillot, L. Corouge, B. Gibaud, G. Le Goualher, D.L. Collins, A. Evans, G. Malandain, N. Ayache, G.E. Christensen, and H.J. Johnson. Retrospective evaluation of inter-subject brain registration. *IEEE Transactions on Medical Imaging*, 22(9):1120–1130, 2003.

[23] Pierre Jannin, J. Michael Fitzpatrick, David J. Hawkes, Xavier Pennec, Ramin Shahidi, and Michael W. Vannier. Validation of medical image processing in image-guided therapy. *IEEE Transactions on Medical Imaging*, pages 1445–1449, 2002.

[24] H.J. Johnson. Method for consistent linear-elastic medical image registration. Master's thesis, Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, IA 52242, May 2000.

[25] H.J. Johnson and G.E. Christensen. Consistent landmark and intensity-based image registration. *IEEE Trans. Med. Imaging*, 21(5):450–461, 2002.

[26] Arno Klein, Jesper Andersson, Babak A. Ardekani, John Ashburner, Brian Avants, Ming-Chang Chiang, Gary E. Christensen, D. Louis Collins, James Gee, Pierre Hellier, Joo Hyun Song, Mark Jenkinson, Claude Lepage, Daniel Rueckert, Paul Thompson, Tom Vercauteren, Roger P. Woods, J. John Mann, and Ramin V. Parsey. Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *NeuroImage*, 46:786–802, 2009.

[27] Luis Lbanez, Will Schroeder, Lydia Ng, Josh Cates, and the Insight Software Consortium. *ITK Software Guide*. Kitware, Inc., New York, 2005.

[28] Weiguo Lu, Ming-Li Chen, Olivera G H, Ruchala K J, and Mackie T R. Fast free-form deformable registration via calculus of variations. *Phys Med Biol*, 49:4621–4637, 2004.

[29] D.S. Marcus, T.R. Olsen, M. Ramaratnam, and R.L. Buckner. The extensible neuroimaging archive toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics*, 5:11–34, 2007.

[30] S. Ogawa, T.M. Lee, A.S. Nayak, and P Glynn. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *agn Reson Med*, 14:68–78, 1990.

[31] M.I. Sereno, A.M. Dale, J.B. Reppas, K.K. Kwong, J.W. Belliveau, T.J. Brady, B.R. Rosen, and R.B.H. Tootell. Borders of multiple visual areas revealed by functional magnetic resonance imaging. *Science*, 268:889–893, 1995.

[32] C.V. Stewart, Chia ling Tsai, and B. Roysam. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. *IEEE Transactions on Medical Imaging*, pages 1379–1394, 2003.

[33] J.P. Thirion. Fast non-rigid matching of 3d medical images. Report 2547, Institut National De Recherche en Informatique Et En Automatique, May 1995.

[34] J.P. Thirion. Image matching as a diffusion process: an analogy with maxwell's demons. *Medical Image Analysis*, 2:243–260, 1998.

[35] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Non-parametric diffeomorphic image registration with demons algorithm. In *MICCAI*, pages 319–326, 2007.

[36] He Wang, Lei Dong, Jennifer ODaniel, Radhe Mohan, Adam S Garden, K Kian Ang, Deborah A Kuban, Mark Bonnen, Joe Y Chang2, and Rex Cheung. Validation of an accelerated demons algorithm for deformable image registration in radiation therapy. *Phys Med Biol*, 50:2887–2905, 2005.

[37] Jay West, J. Michael Fitzpatrick, et al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *J. Comp. Asst. Tomog.*, 21(4):554–566, 1997.

[38] R.P. Woods, S.T Grafton, J.D. Watson, N.L. Sicotte, and J.C. Mazziotta. Automated Image Registration: II. Intersubject Validation of Linear and Nonlinear Models. *Journal of Computer Assisted Tomography*, 22(1):153–165, 1998.

[39] M.A. Yassa and C.E.L. Stark. A quantitative evaluation of cross-participant registration techniques for mri studies of the medial temporal lobe. *NeuroImage*, 44:319–327, 2009.

[40] Barbara Zitova and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000, 2003.