Report

Medical Image Registration with Information Theory

Yuwei Qiu

Working diary 1

Conduct a literature survey in medical image registration, and answer the following questions:

(a) What is the scientific goal of medical image registration?

(b) What are the most commonly used methods in medical image registration?

(c)How can we compare their relative advantages and disadvantages?

Medical Image Registration: an overview and basic methodologies

1 Introduction

Over the past decades, medical imaging technologies have evolved from 2D projections to sophisticated 3D digital modalities. 2D imaging technologies involve X-ray projections for example and 3D methods are far more dimensional and vivid like computed topography (CT) and magnetic resonance imaging (MRI). Moreover, specific imaging technologies build connection between projection and volume, single photo emission computed topography (SPECT) for instance. The revolutionary breakthrough of information technologies provide great chances to evaluate messages conveyed by these images more accurate and effectively.

These different physical projections or modalities deliver valuable messages (electron density, magnetic behavior of hydrogen nuclei, reflection of ultrasound from tissue interfaces, etc.).[1] During the diagnosis, doctors tend to refer to various messages of a patient attained through medical images. However, for these messages are classified into different datasets, it is quite inconvenient to consolidate individual pieces into a single narrative.

Therefore, it is fundamental before the information makes sense to for the information to correspond. The process develops into a significant branch of image signal processing.

What doctors need is definitely not different datasets telling their own stories. They need a far more coherent analysis or a single data space. This is where medical image registration plays a crucial role.

Medical Image Registration refers to the implement of establish correspondences in two or more images, pixel to pixel or voxel to voxel. It is now playing an increasingly important role in many clinical applications including Computer-assisted Diagnosis (CAD). [2]

This passage will discuss over several basic issues of medical image registration including the target, the methodologies and evaluation of each method and the criteria of evaluating a registration method.

2 The goal of medical image registration

To make a detailed explanation, several examples are raised below. Medical Image Registration can be applied to images from the same subject acquired by different imaging modalities or projections as well as at different time points, which refers to multi-modal image registration and serial image registration. These two kinds of registration could assistant in the evaluation of a specific subject from spatial construction or long lasting development. Also, another application is inter-subject registration, aiming to align images acquired from different subjects through population.

After proper and valid registration, datasets could be transformed into a single and coherent database, which serves as important references in diagnosis.

3 Methodologies on Medical Image Registration

As discussed before, different images present various messages. Also, a single image could be different extractions of picture characteristics including points, curves and surfaces etc. [3] Such methods analyzing specific points or points clouds discrete called Point Method and Feature Method.

As we conduct enough registration, the images information will build up a templates database. By carefully matching using mathematical tools including matrix or Fourier Transform, we shall figure out the correlation between the images we want to register and the templates, which is Correlation and Sequential Method.

3.1 Point Method (based on the extraction of points)

Registration begins with the correspondence of particular points and when specific points are matched, the registration is complete.
In images registration, knowledge about corresponding points in two images

is required to spatially align the images. It is important that detected points be independent of noise, blurring, contrast, and geometric changes so that the same points can obtained in images of the same scene taken under different environmental conditions and sensor parameters. [2]

3.1.1 Global Method and Its Evaluation

Global method based on point matching use a set of matched points to generate a single optimal transformation. By derive the parameters of transformation through approximation, image registration is complete. [4]
Since a large number points are selected for proper approximation, the computing complexity is a problem. And using the normal equations to solve the least squares approximation becomes unstable and inaccurate. Therefore, we introduce the terms of orthogonal polynomials as the terms of polynomial mapping. However, the major limitation of the global method is that this transformation cannot account for local geometric distortions. [4]

3.1.2 Local Method and Its Evaluation

According to different local piece or neighborhoods, multiple computations are performed in local method by using equal control points selected in Global Method.
In local method, a spatial mapping transformation for each coordinate is specified which interpolates between the matched coordinate values. [4]
Since only a part of the selection of points are involved in computation, the capacity and time complexity is declined in local method. However, some local computation of parameters are virtually inherent from global transformation. Actually, on lower resolution datasets, global methods precede their use.

3.1.3 Evaluation of point method

Point method is the fundamental and the simplest one among enormous registration methods. Features of images are reflected from points clouds then analyzed by software pixel by pixel. For points are the basic components of medical images, the error caused by transformation is avoided. And the points clouds are chosen randomly, which leads to statistically logical. [2]
On the contrary, precisely because the selection of the points are randomly, some extreme circumstance might affect the experiments results. For example, if we register two cancers' MRI images, the precision of the devices, the procedures and even the definition of certain images could result in errors or extreme points. In the lack of specific transformation or logical approximation, researches should arrive at wrong conclusion.

3.2 Features methods (based on the features of images)

Image features provide a critical source of information for various recognition tasks. A feature may measure a global or a local image property, revealing statistical, algebraic, geometric, spatial, differential, and spectral information about an image.

3.2.1 Curved method (based on the extractions of curves)

Especially for 2-D projections, this method is quite equivalent and it is virtually an expansion of 3.1.1.

By fitting the curvature of open curves taken from different projections or modalities, a best suitable part of the curved lines is selected. Then discrete the line to specific points and analyze the correspondence of each points cloud.

3.2.2 Surface method (based on the extraction of faces)

One of the most famous algorithm is Head and Hat Method. The Surface taken from one dataset is called Head and the outline taken from the other dataset is Hat. By analyzing the root-mean-square value between each point and fine adjusting the Head and Hat, the registration could be complete.

3.3 Moment and Principal Method and Its Evaluation

The analytic signature is a recently proposed 2D shape representation scheme. It is tailored to the representation of shapes described by arbitrary sets of unlabeled points, or landmarks, because its most distinctive feature is the maximal invariance to a permutation of those points. The shape similarity of two point clouds can then be obtained from a direct comparison of their representations.

Features that are stable under noise and invariant to changes in geometry and contrast of an image are most useful. Invariant features enable comparison of images taken from different views, under different lighting conditions, and with different sensors. Therefore, a mass of data is available when register two images.

However, that is also where goes troublesome. Among the many types of features that can be extracted from an image, one is faced with the problem of selecting those features that carry the most information about the image.

3.4 Correlation and Sequential Method (Similarity method)

3.4.1 Correlation Method and Its Evaluation

The phase correlation method has been used to estimate the translation and rotation between similar images on the 2-D plane in similarity method. It is conducted by mathematic tools like matrices in linear algebra. By the convolution theorem, correlation can be computed as a product of Fourier transforms. [4]

The use of Fast Fourier Transform (FFT) makes the process more efficiently. [4] On the contrary, there are three major caveats. Treated by FFT, it requires a memory capacity that grows with O(logN). Also, only the cross-correlation before normalization could be implemented by FFT. Moreover, Correlation Measures that are sensitive to radiometric changes in the scene or invariant to sensor parameters, which leads to the unclear analysis of the noisy image. [4]

3.4.2 Sequential Method and Its Evaluation

The Sequential Method was proposed by Barnea, which is a revolution of Correlation Method. They suggest a similarity measure instead, which is computationally much simpler.

The Sequential Method can significantly reduce the complexity with minimal performance degradation. However, as the degrees of freedom of the transformation is increased, they still have increasing complexity. [4]

3.4.3 Evaluation of Similarity Method

Various similarity measures have been formulated throughout the years, each with its own strengths and weaknesses. Some measures use raw image intensities, some normalize the intensities before using them, some use the ranks of the intensities, and some use joint probabilities of corresponding intensities.

There are also problems with similarity method. In one problem, an observed image and a number of saved images are given and it is required to determine the saved image that best matches the observed image. The second problem involves locating an object of interest in an observed image where the model of the object is given in the form of a template and the observed image is an image being viewed by a camera. To locate the object within the observed image, there is a need to find the best-match position. [2]

3.4 Maximization of Mutual Information

3.4.1 Mutual Information

In probability theory and information theory, the mutual information (MI) or (formerly) transformation of two random variables is a measure of the variables' mutual dependence. Not limited to real-valued random variables like the correlation coefficient, MI is more general and determines how similar the joint distribution p(X,Y) is to the products of factored marginal distribution p(X)p(Y). [1]

3.4.2 Maximization of Mutual Information

A new approach to the problem of multimodality medical image registration is proposed, using a basic concept from information theory, or relative entropy, as a new matching criterion. This method measures the statistical dependence or information redundancy between the image intensities of corresponding voxels in both images, which is assumed to be maximal if the images are geometrically aligned. [1]

3.4.3 Evaluation of Maximization of Mutual Information
Maximization of MI is a very general and powerful criterion, because no assumptions are made regarding the nature of this dependence and no limiting constraints are imposed on the image content of the modalities involved. The accuracy of the MI criterion is validated for rigid body registration of computed tomography (CT), magnetic resonance (MR), and photon emission tomography (PET) images by comparison with the stereotactic registration solution, while robustness is evaluated with respect to implementation issues, such as interpolation and optimization, and image content, including partial overlap and image degradation. [1]

4 The Standard of Evaluation Registration Methods

Various methods could be the best registration application according to different tasks. The characteristics including feature space, similarity metric and search space and strategy are major components of registration methods. By detailed evaluation of these three components, we could evaluate the complexity.
Meanwhile, the complexity of registration methods is an important criterion when determine whether the method is suitable. The complexity of time or capacity are determined by the transformation classes, which are influenced by the source of misregistration.[4]

References

[1] Ayman S.EI-Baz, Rajendra AcharyaU, Andrew F.Laine. Medical Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies [M]. ISBN 978-1-4419-8203-2. London: Springer Science+Business Media, 2011,

[2]Thomas M. Deserno. Biochemical Image Progressing [M]. ISBN 978-3-642-15815-5. Berlin: Springer-Verlag Berlin Heidelberg, 2011.

[3]Nikos Paragios, Yunmei Chen, Oliver Fargeras. Handbook of Mathematical Models in Computer Vision[M]. ISBN 0-387-26371-3 USA: Springer Science+Business Media, 2006.

[4]Lisa Gottesfeld Brown. A Survey of Image Registration Techniques. [N] New York: Columbia University, 1992.

Working diary 2

Use a programming language to experiment with medical image registration with various methods.

Medical Image Registration in MATLAB: Case Study

Abstract: MATLAB is the most prevalent programming environment in image processing. Based on MATLAB functions, this passage gives an ordinary process of Medical Image Registration.

Key Words: MATLAB. Medical Image Registration

Ⅰ Passage Map and Introduction

1 Methods discussed

Based on the literature review, several methods are brought forward. This passage selects a few, which are listed behind, and discusses the implement on MATLAB in a case-solving order.

Methods include:
- Point Map Methods
- Features Methods (based on Point Mapping)
- Similarity Transformation Method
- Method of Mutual Information
- Rigid Matrix Transformation Method

2 General Order of Medical Image Registration

Procedures	Remarks
Classify Images	Different categories of medical images are uploaded into database. We usually assort rigidly before we include the imaged into the platform.
Compare Images	Use specific function to compare two

	different images.
Establish Criteria	
Display Images and Rough Registration	Through especial registration methods.
Accurate Registration	Slightly change the step length and use iteration to calibrate the images.
Examination of Registration	Set specific standard of the result of medical image registration and quit iteration.

3 Programming Environment

This passage mainly discuss the implement of Medical Image Registration in the programming environment of MATLAB.

MATLAB (MATrix LABoratory) is a multi-paradigm numerical computing environment and fourth-generation programming language. A proprietary programming language developed by MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, Java, Fortran and Python.

Ⅱ Case study

In this part of passage, I will carry out a complete registration towards two of the images

1 Display

In order to study the process thoroughly, I downloaded a zip of files including several medical images from the website of DICOM Information Resources.

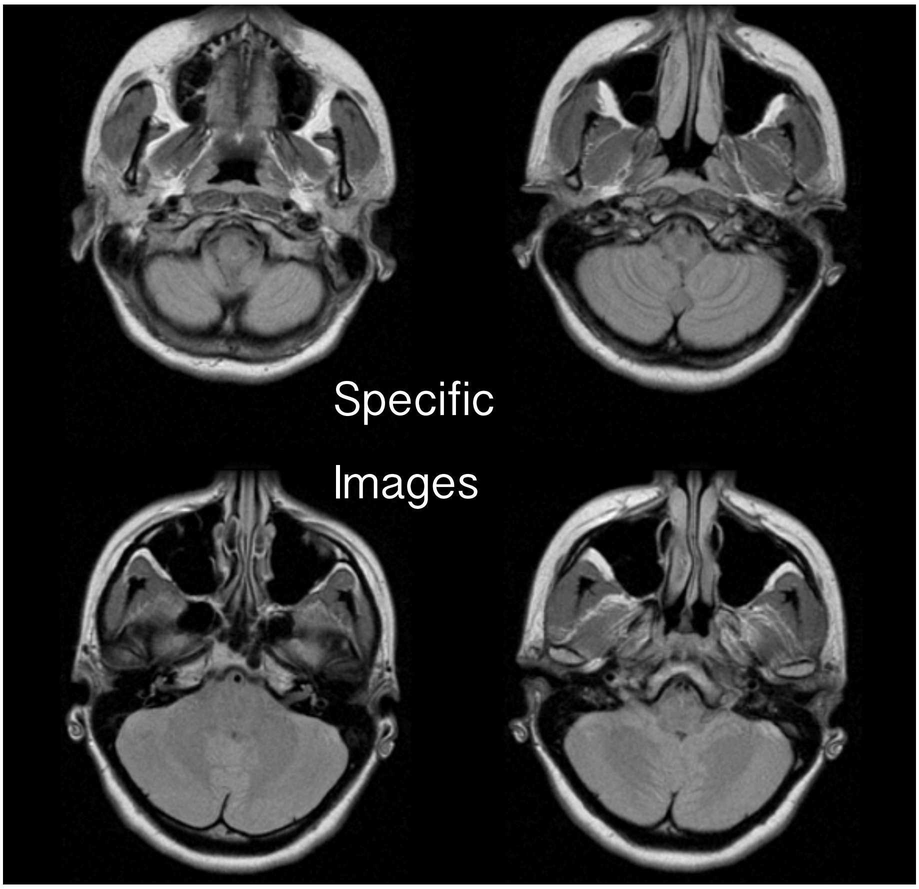
Some of the images are attached below. [Figure 2-1-1]

For the formats of these kind of images, dcm, is an special sort of MATLAB sources, which could not be displayed as ordinary picture formats like bmp, jpg, tiff, gif, pcx, tga, exif, fpx, svg, psd, cdr, pcd, dxf, ufo, eps, ai, raw.

In this form [Form 2-1-1] as well as the forms in the following passage, both of the structure of the code and the sources code are written in the "Code".

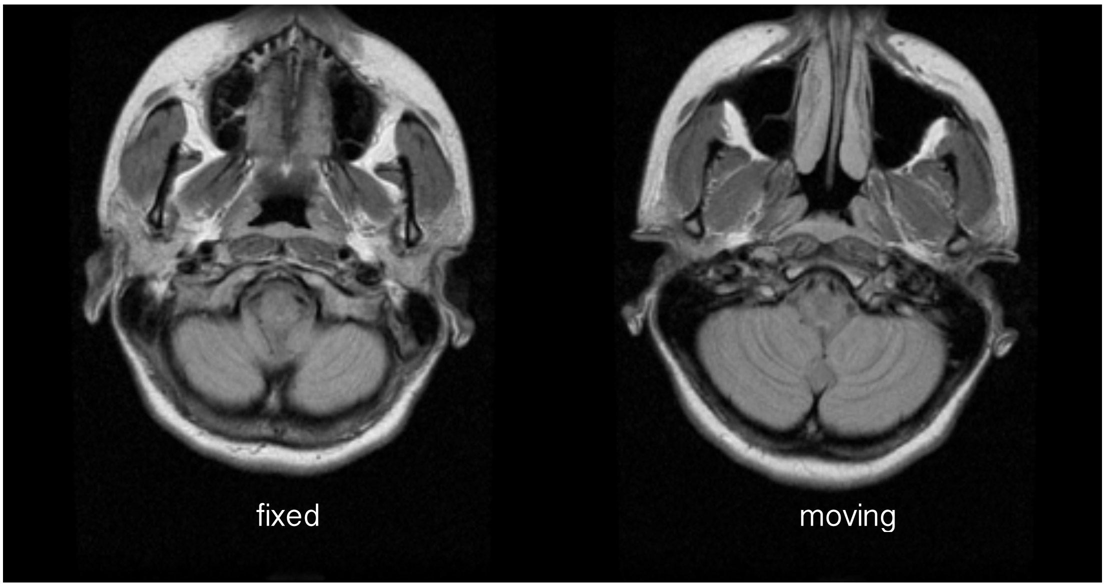
Code	Remarks	Function
>>a1=imread('address'); >>a1=imread(' C:\Users\victoria\Desktop\digest_article\brain_003');	Read files Address refers to the address of specific dcm image	A = imread(filename, fmt) reads a grayscale or color image from the file specified by the string filename. If the file is not in the current folder, or in a folder on the MATLAB® path, specify the full pathname.
>> figure, imshow(img, 'DisplayRange',[]); >> figure, imshow(a1, 'DisplayRange',[]);	Show files	imshow(X,RX,map) displays the indexed image X with associated 2-D spatial referencing object RX and colormap MAP.

[Form 2-1-1]



[Figure 2-1-1]

In the registration, I select the following two images as the fixed images and the moving images [Figure 2-1-2]:



[Figure 2-1-2]

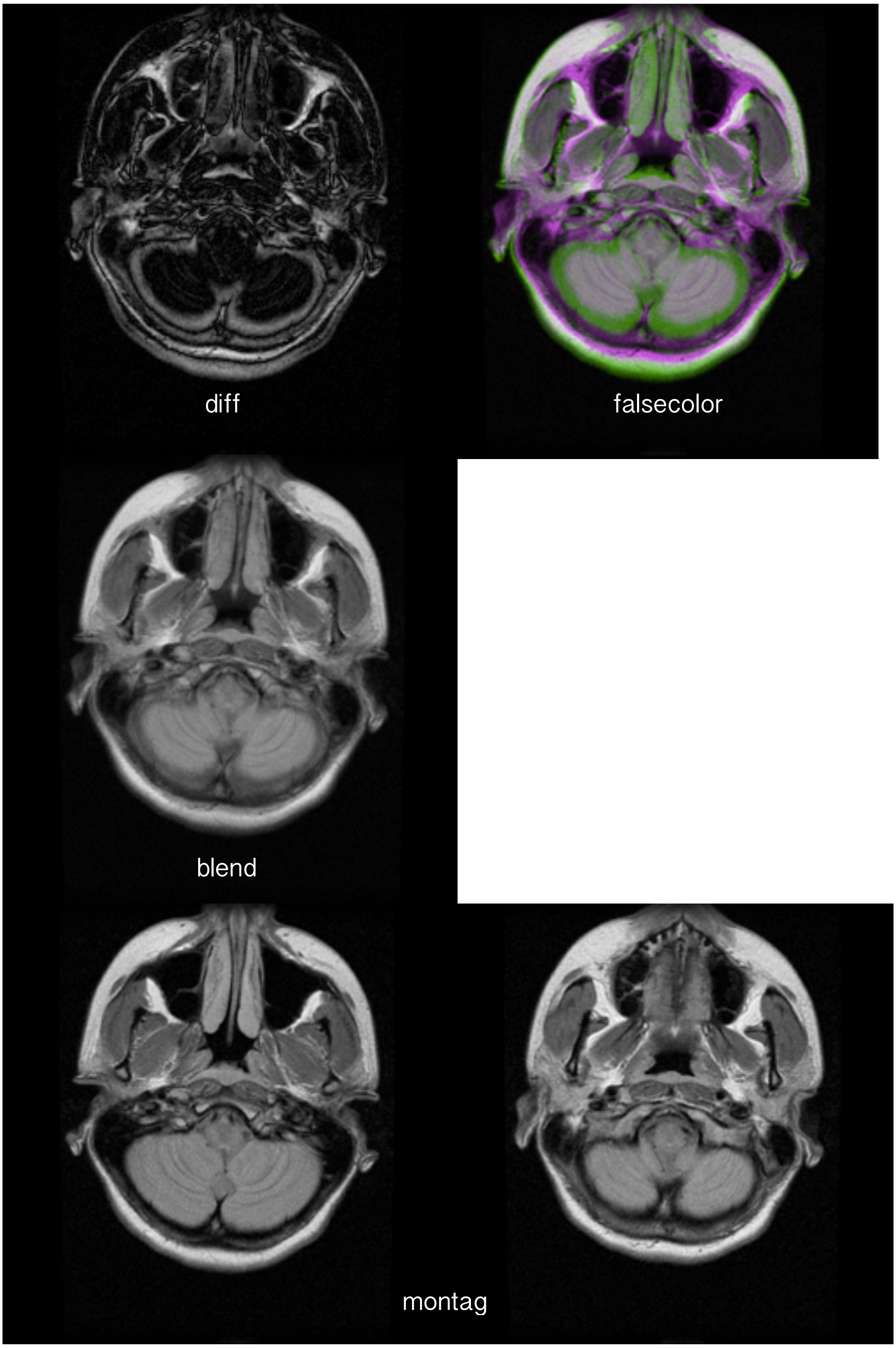
2 Comparison

Before we carry out the registration, we need to attain a general recognition of the difference between two specific images.

Code	Remarks	Function
>> figure = imshowpair(moving, fixed, 'method'); >> figure = imshowpair(moving, fixed, 'blend'); >> figure = imshowpair(moving, fixed, 'falsecolor'); >> figure = imshowpair(moving, fixed, 'diff'); >> figure = imshowpair(moving, fixed, 'montage');	Compare the images and display them	imshowpair(img2, img1, 'method'); method: diff, blend, falsecolor and montage. Display a pair of grayscale images with two different visualization methods, 'diff' and 'blend'. Display the images with the default method 'falsecolor' and apply brightness scaling independently to each image.

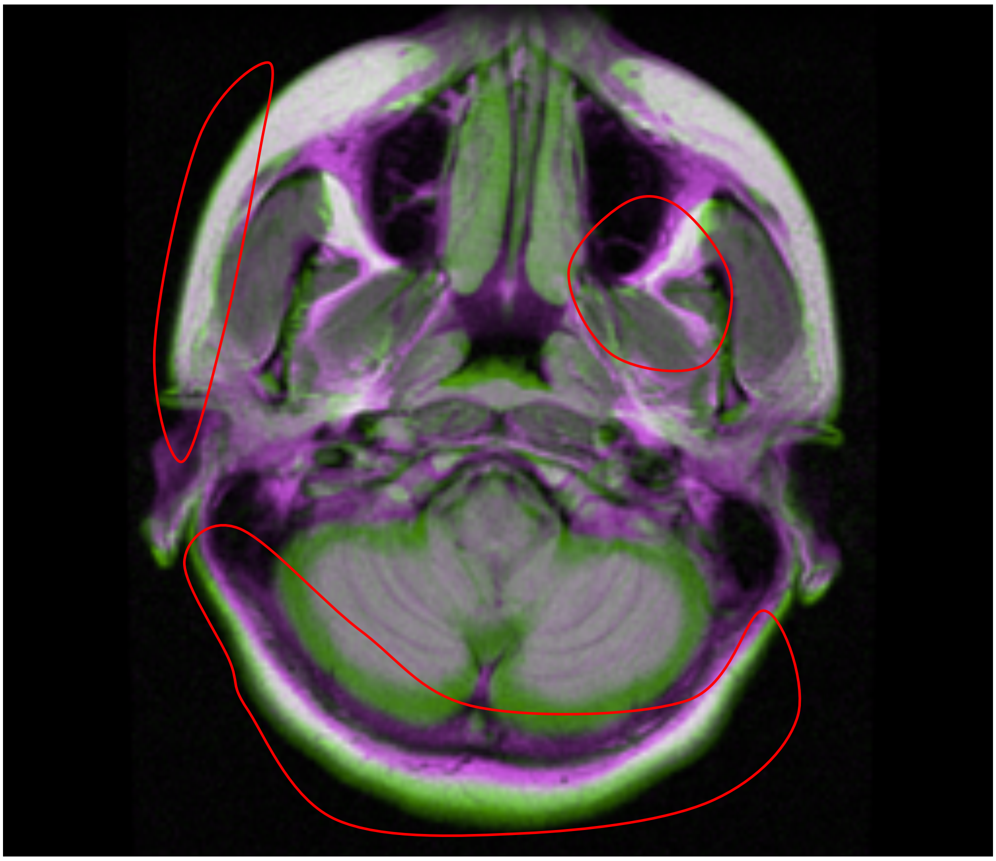
[Form 2-2-1]

Different methods of comparison depict various characteristics of images. Several images of the comparison are listed in the next page.[Figure 2-2-1]



[Figure 2-2-1]

From the unregistered image we could see there are few blurred areas in the image:
These areas are circled behind. Take the blend image as an instance.[Figure 2-2-2]



[Figure 2-2-2]

Then we carry out two procedures of registration, rigidly and accurately.

3 Criteria

This part of registration is related to the general registration reviewed in Task 1.
We commit several transformation method like Mean Square method and Mutual Information method.

Fortunately, we do not have to regard the complex mathematical process and equations. MATLAB offers several functions already using these transformation method.

Code	Remark
>>[optmizer, metric] =	Methods refer to 'monomodal'

imregconfig('modality'); >>[registration.optimizer.RegularStepGradientDescent, registration.metric.MeanSquares] = imregconfig('monomodal')	and 'multimodal'. Optimizers refer to registration. optimizer. RegularStepGradientDescent and registration. optimizer. OnePlusOneEvolutionary. Metrc refers to different transformation method including registration.metric.MeanSquare and registration.metric.MattesMutualInformation.

[Figure 2-3-1]

4 Rough Registration

By starting rigid registration, we finally begin to register the images using the criteria we established in the previous section.

Code	Remarks	Function
>> movingRegisteredDefault = imregister(moving, fixed, 'translation', optimizer, metric); >> movingRegisteredDefault = imregister(moving, fixed, 'translation', registration.optimizer.RegularStepGradientDescent, registration.metric.MeanSquares); >> movingRegisteredDefault = imregister(moving, fixed, 'rigid', registration.optimizer.RegularStepGradientDescent, registration.metric.MeanSquares); >> movingRegisteredDefault = imregister(moving, fixed, 'similarity', registration.optimizer.RegularStepGradientDescent, registration.metric.MeanSquares); >> movingRegisteredDefault = imregister(moving, fixed, 'affine', registration.optimizer.RegularStepGradientDescent, registration.metric.MeanSquares);	The optimizer and metric must be coherent with the ones appeared in the criteria establishment.	moving_reg = imregister(moving, fixed, transformType, optimizer, metric) transforms the 2-D or 3-D image, moving, so that it is registered with the reference image, fixed. Both moving and fixed images must be of the same dimensionality, either 2-D or 3-D. transformType is a character string that defines the type of transformation to perform. optimizer is an object that describes the method for optimizing the metric and metric is

		an object that defines the quantitative measure of similarity between the images to optimize. Returns the aligned image, moving_reg.

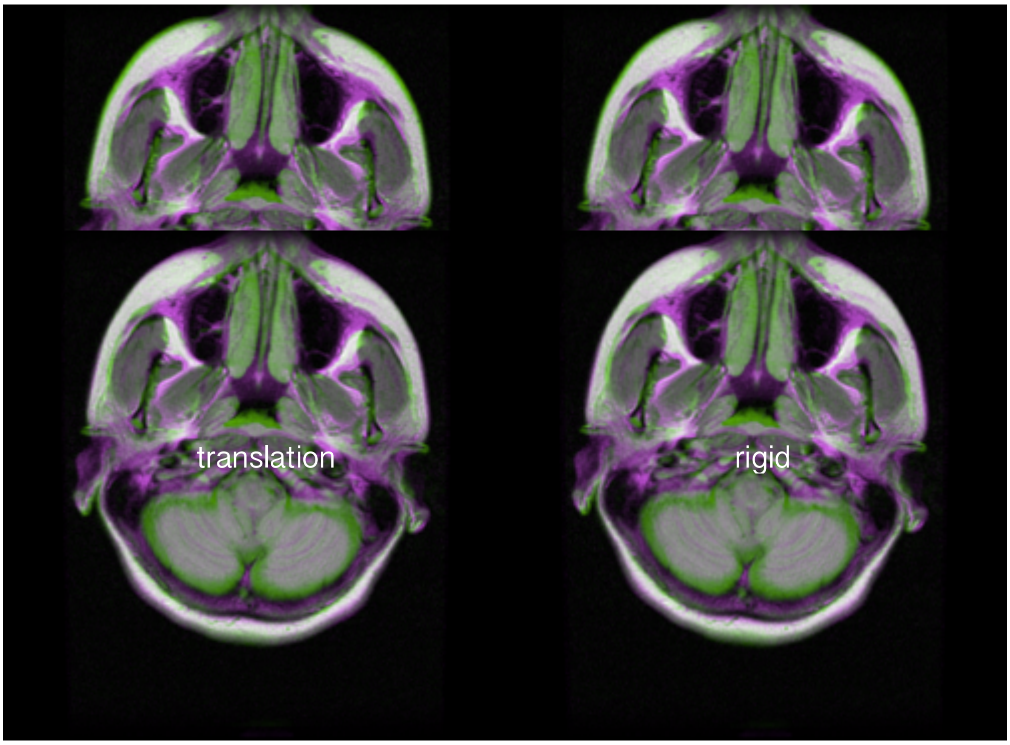
[Figure 2-3-2]

In this part, I commit four kinds of transformation in order to compare their effectiveness and accuracy. Results are listed in the pictures below.[Figure 2-4-1]

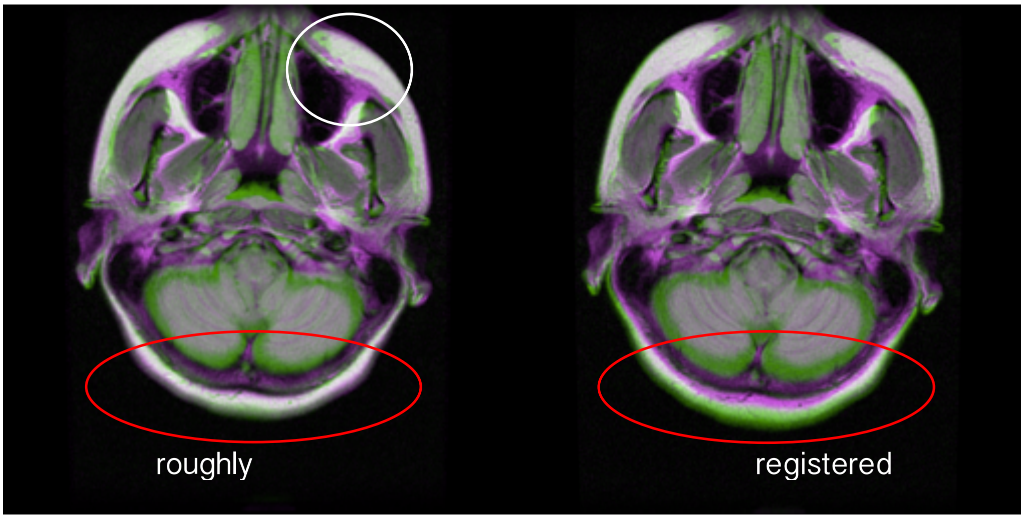
From rough observation, we can hardly figure out the difference among the four kinds of transformations from the results they give.

Firstly, the images are quite simple for they are attained through the same medical imaging method (computer topography). So after an simple transformation, they are alike with each other in the database.

Secondly, though the blurred areas narrow down and the numbers of shadowed regions declined, there are still some places which is not perfectly registered, the right up-corner for example. [Figure 2-4-2]



[Figure 2-4-1]



[Figure 2-4-2]

5 Modify the Optimizer and Accurate Registration

In the diagnosis, images should depict information more accurately.

The registration is virtually iteration theoretically, therefore we could slightly change the step length and times of iteration to attain a more precise result.

Firstly, we must recognize the previous step length.

Tapping “>> disp(registration.optimizer.RegularStepGradientDescent)”, we get the following description:

```
>> disp(registration.optimizer.RegularStepGradientDescent)
  registration.optimizer.RegularStepGradientDescent

  Properties:
    GradientMagnitudeTolerance:  1.000000e-04
             MinimumStepLength:  1.000000e-05
             MaximumStepLength:  6.250000e-02
             MaximumIterations:  100
               RelaxationFactor:  5.000000e-01
```

By using specific function “disp()”, we could change the properties of optimizer we selected previously.

Code:

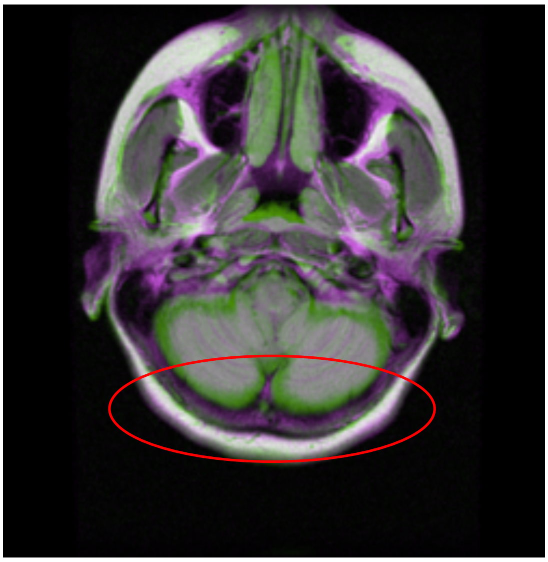
```
>>registration.optimizer.RegularStepGradientDescent.MinimumStepLength
=
registration.optimizer.RegularStepGradientDescent.MinimumStepLength/3.
5;
                                        % divide the Step length by 3.5
>>registration.optimizer.RegularStepGradientDescent.MaximumIterations =
registration.optimizer.RegularStepGradientDescent.MaximumIterations*10
0;
                                        % multiply the iteration by 100
>>movingRegisteredAdjustedStepLengthTimes = imregister(moving, fixed,
'affine',            registration.optimizer.RegularStepGradientDescent,
registration.metric.MeanSquares);
                              % register the image with renewed optimizer
>> figure, imshowpair(movingRegisteredAdjustedStepLengthTimes, fixed);
                                                    % display the figure
```

After the change, the properties of the optimizer are modified:

```
>> disp(registration.optimizer.RegularStepGradientDescent)
  registration.optimizer.RegularStepGradientDescent

  Properties:
    GradientMagnitudeTolerance: 1.000000e-04
             MinimumStepLength: 2.857143e-07
             MaximumStepLength: 6.250000e-02
             MaximumIterations: 100000000
               RelaxationFactor: 5.000000e-01
```

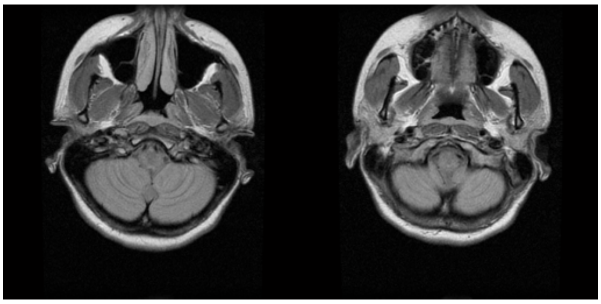
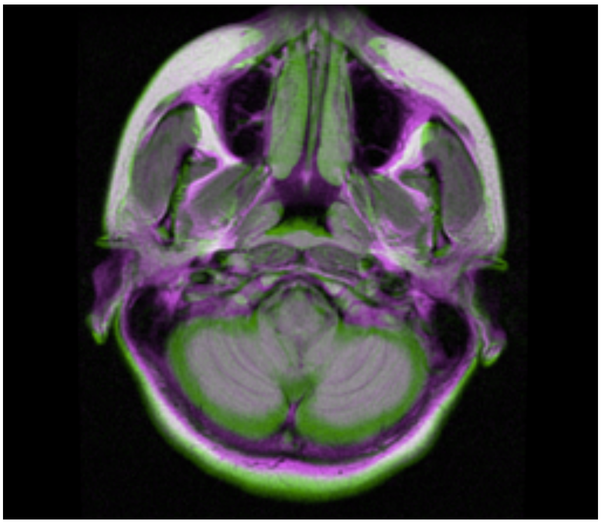
The result we get is more accurate and the blurred areas are narrow down a bit more.

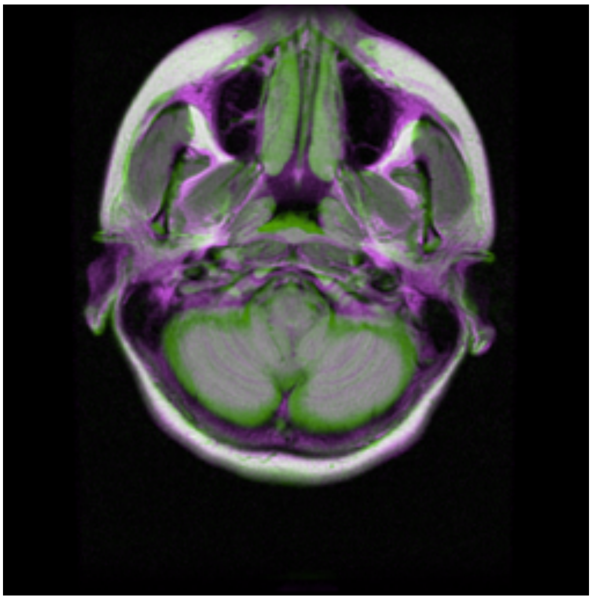


Figure[2-5-1]

III Summary

In this case, we generally register two images of human brain attained from CT. The following format gives a brief summary of the process and every procedure.

Sections	Remarks	Result
display	General display of solid images attained from medical imaging methods	
establish criteria	Select proper optimizer and metric	`>> disp(registration.optimizer.RegularStepGradientDescent)` registration.optimizer.RegularStepGradientDescent Properties: GradientMagnitudeTolerance: 1.000000e-04 MinimumStepLength: 1.000000e-05 MaximumStepLength: 6.250000e-02 MaximumIterations: 100 RelaxationFactor: 5.000000e-01
rough registration	First registration (Blurred)	
Change criteria	Divide step length Multiply iterations	`>> disp(registration.optimizer.RegularStepGradientDescent)` registration.optimizer.RegularStepGradientDescent Properties: GradientMagnitudeTolerance: 1.000000e-04 MinimumStepLength: 2.857143e-07 MaximumStepLength: 6.250000e-02 MaximumIterations: 100000000 RelaxationFactor: 5.000000e-01

accurate registratio n	Second Registration Divided step length Multiplied iterations	

The registration process is complete.

Moreover, different images required different method of transformations, which leads to different methods of rough and accurate registration for it causes different functions used in the process. However, in this passage, for a lot more images and statistics are required in the analysis to gain a more specific result of evaluation, little analysis over different methods are discussed.

Source code

```
%initial the two images
>>
a1=dicomread('C:\Users\victoria\Desktop\digest_article\brain_003');
>> a2=dicomread('C:\Users\victoria\Desktop\digest_article\brain_004');
>> fixed=a1;
>> moving=a2;

%display the unregistered image
>> figure = imshowpair(moving, fixed, 'falsecolor');

%establish the criteria
>> [registration.optimizer.RegularStepGradientDescent,
registration.metric.MeanSquares] = imregconfig('monomodal')

%rought registration
>> movingRegisteredDefault = imregister(moving, fixed, 'translation',
registration.optimizer.RegularStepGradientDescent,
registration.metric.MeanSquares);
>> figure, imshowpair(movingRegisteredDefault, fixed);

%change the properties of optimizer
>> disp(registration.optimizer.RegularStepGradientDescent)
>>registration.optimizer.RegularStepGradientDescent.MinimumStepLength
=
registration.optimizer.RegularStepGradientDescent.MinimumStepLength/3.
5;
>>registration.optimizer.RegularStepGradientDescent.MaximumIterations =
registration.optimizer.RegularStepGradientDescent.MaximumIterations*10
0;

%accurate registration
movingRegisteredAdjustedStepLengthTimes = imregister(moving, fixed,
'affine', registration.optimizer.RegularStepGradientDescent,
registration.metric.MeanSquares);

%depict the final result
>> figure, imshowpair(movingRegisteredAdjustedStepLengthTimes, fixed);
```

Reference

1. Gonzalez. Digital Image Processing (Second Edition)[M]. ISBN 978-3-642-15815-5. Berlin: Springer-Verlag Berlin Heidelberg, 2011.
2. MATLAB functions descriptions

Working diary 3

The next stage is to think about how to improve the performance of existing medical image registration methods.

Essentially, this line of work shows that the traditional approach of using the Maximum Likelihood Estimator (MLE) in entropy estimation is highly sub-optimal in the large alphabet regime, and replacing the MLE with ones that are more theoretically sound may lead to significant performance boosts in various practical tasks.

Mutual information based methods in medical image registration have received significant attention in past years, and we refer the project participants to, e.g. [2] for a review.

In this week, I experiment with the mutual information based medical image registration method, compare it with week 1-2, and write a comprehensive report.

Maximum Likelihood Estimator

and

JVHW Estimator

Review, Application, and Comparison

Ⅰ Introduction

In probability theory and information theory, the mutual information (MI) or (formerly) transformation of two random variables is a measure of the variables' mutual dependence. Not limited to real-valued random variables like the correlation coefficient, MI is more general and determines how similar the joint distribution p(X,Y) is to the products of factored marginal distribution p(X)p(Y).

A new approach to the problem of multimodality medical image registration is proposed, using a basic concept from information theory, or relative entropy, as a new matching criterion. This method measures the statistical dependence or information redundancy between the image intensities of corresponding voxels in both images, which is assumed to be maximal if the images are geometrically aligned.

Maximization of MI is a very general and powerful criterion, because no assumptions are made regarding the nature of this dependence and no limiting constraints are imposed on the image content of the modalities involved. The accuracy of the MI criterion is validated for rigid body registration of computed tomography (CT), magnetic resonance (MR), and photon emission tomography (PET) images by comparison with the stereotactic registration solution, while robustness is evaluated with respect to implementation issues, such as interpolation and optimization, and image content, including partial overlap and image degradation. [1]

Ⅱ Maximum Likelihood Estimator

2.1 Introduction

Maximum Likelihood Estimation is a kind of plug-in method. Plug-in refers to a process of seeking for the best smooth curved faces of n-D to fit a group of given points. Maximum Likelihood Estimator is a statistical method of estimating the parameters of a statistical model with given data.

To illustrate, Maximum Likelihood Estimator is a kind of output of Maximum Likelihood Estimation. Suppose in a random experiment with the results of A, B, C and D. Different conditions leads to distinct probability of the results in a single experiment. Therefore, in one specific random experiment with the result of A, we might assume that the probability of A is higher than any one of the others due to the inadequate times of experiments. The probability of A is related to a parameter theta,θ. One of the main results of Maximum Likelihood Estimator is to give the functional equation of such parameters.

Maximum-likelihood estimation gives a unified approach to estimation, which is well-defined in the case of the normal distribution and many other problems. However, in some complicated problems, difficulties do occur: in such problems, maximum-likelihood estimators are unsuitable or do not exist.

2.2 Evaluation of MLE

2.2.1 Finding the function of theta

To evaluate the result and complexity of MLE, we need to figure out the function of the estimator. In a word, we have to give a function with a single variable of theta. One general method to estimate unknown parameters is to find the function of the statistical data points and the parameter.

Suppose there is a group of samples p1, p2, …, pn of n independent and identically distributed observations, coming from a distribution with an unknown probability density function $f(p)$
If we fixed the samples, theta is variable in such circumstances. We define the function of theta as $L(\theta)$, therefore we come to a fundamental equation with joint density function

$$L(\theta)=\prod_{i=1}^{n} f(p_i)$$

With both sides in logarithm then divided by n, we have the average of $L(\theta)$, which we defined as $\hat{l}(\theta)$:

$$\hat{l}(\theta) = \frac{1}{n}\ln L(\theta) = \frac{1}{n}\ln\left(\prod_{i=1}^{n} f(p_i)\right) = \frac{1}{n}\sum_{i=1}^{n} f(p_i|\theta)$$

Here we use a vector $P = [\, p_1, p_2, \dots \dots p_i \,]$ to represent the sample and we define $F(P) = \sum_{i=1}^{n} f(p_i)$

2.2.2 Entropy

Generally, entropy, more specifically Shannon entropy, refers to disorder or uncertainty. In information theory, systems are modeled by a transmitter, channel, and receiver. The transmitter produces messages that are sent through the channel. The channel modifies the message in some way. The receiver attempts to infer which message was sent. Entropy is the expected value (average) of the information contained in each message. "Messages" can be modeled by any flow of information.
Entropy is defined by the following function:

$$H(P) \triangleq \sum_{i=1}^{S} -p_i \ln p_i,$$

Next, we could analyze the risk of MLE. Here we define $R(P)$ as the risk of MLE.
According to [1], suppose we have a sample with unknown size of S, risk of MLE could be written as [2]

$$\sup_{P \in \mathcal{M}_S} \mathbb{E}_P (H(P_n) - H(P))^2 \asymp \frac{S^2}{n^2} + \frac{(\ln S)^2}{n}$$

With another research [3], for any fixed distribution P of S elements, we have

$$\mathbb{E}_P\left(-P_n(i)\ln P_n(i)\right) = -p_i \ln p_i - \frac{1-p_i}{2n} + O\left(\frac{1}{n^2}\right)$$

With the risk of $O(\frac{1}{n^2})$, $n \gg \frac{S}{\ln S}$ is both necessary and sufficient, which is quite unfriendly in statistical estimation.

The Shannon entropy can also be computed for an image, in which case we do not focus on the probabilities of letters or words occurring, but on the distribution of the gray values of the image. A probability distribution of gray values can be esti- mated by counting the number of times each gray value occurs in the image and dividing those numbers by the total number of occurrences. An image consisting of almost a single intensity.

Ⅲ Medical Image Registration (MIR) based on Mutual Information

1 Introduction

The research that eventually led to the introduction of mutual information as a registration measure dates back to the early 1990s. Woods et al. [4], [5] first introduced a registration mea- sure for multimodality images based on the assumption that regions of similar tissue (and, hence, similar gray values) in one image would correspond to regions in the other image that also consist of similar gray values (though probably different values to those of the first image). Ideally, the ratio of the gray values for all corresponding points in a certain region in either image varies little. Consequently, the average variance of this ratio for all regions is minimized to achieve registration. Here, Fig 1 shows the resulting histograms when one MR image is rotated with respect to the other by angles of 2°, 5° , and 10° . [6]

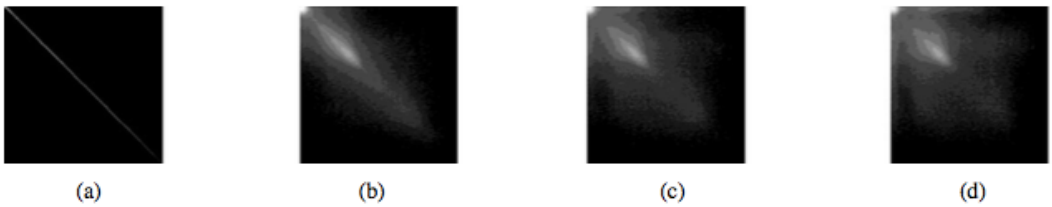


Fig 1 [6]

2 literature review and brief summary on Mutual Information Methodologies of (MIR)

The class “Method” can be further subdivided into preprocessing, measure, transformation, and implementation.

Preprocessing	• Define a region or structures of interest in the images to exclude structures that may negatively influence the registration results. • Low-pass filtering to remove speckle in ultrasound images and thresholding or filtering to remove noise	• Blurring (correct for differences in the intrinsic resolution of the images.) • Resample the images (achieve similar voxel sizes in all image dimensions or obtain similar voxel sizes in the images to be registered)

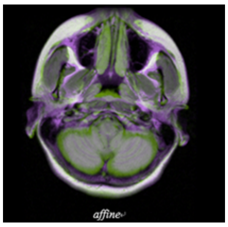
Measure	Entropy (especially Shannon Entropy)	
	Normalization	
	Spatial Information	
Implementation	The importance of the implementation of a mutual information based method should not be underestimated, since implemental decisions can have a large influence on the registration results.	- Interpolation: - Probability Distribution Estimation - Optimization - Acceleration - ……
Transformation	The transformation applied to register the images can be categorized according to the degrees of freedom.	- Rigid - Affine - Curved - …… 

Table 1

Ⅳ MIR Based on Mutual Information

Here in Section 4, I would use Powell method to register two medical images of human brain, which were also the target pictures in the previous report last weeks.

Code:

```
function [mi]=PV(x,y,ang,I,J)    //Powel.m

%       x         bit shift of x-axis
%       y         bit shift of y-axis
%       ang      angle of rotation
%      Powell plug-in method, return the value of joint histogram


a=double(I);
b=double(J);
[M,N]=size(a);
hab=zeros(256,256);
ha=zeros(1,256);
hb=zeros(1,256);
if max(max(a))~=min(min(a))
      a=(a-min(min(a)))/(max(max(a))-min(min(a)));
else
      a=zeros(M,N);
end
if max(max(b))-min(min(b))
      b=(b-min(min(b)))/(max(max(b))-min(min(b)));
else
      b=zeros(M,N);
end

a=double(int16(a*255))+1;
b=double(int16(b*255))+1;
%transformation of reflection

[width,height]=size(b);
u=(width-1)/2;
v=(height-1)/2;
rad=pi/180*ang;
t1=[1 0 0;0 1 0;x y 1];
t2=[1 0 0;0 1 0;-u -v 1];
```

```
t3=[cos(rad) -sin(rad) 0;sin(rad) cos(rad) 0;0 0 1];
t4=[1 0 0;0 1 0;u v 1];
T=t2*t3*t4*t1;
tform=maketform('affine',T);
coordinate_x=zeros(width,height);
coordinate_y=zeros(width,height);
for i=1:width
    for j=1:height
        coordinate_x(i,j)=i;
    end
end
for i=1:width
    for j=1:height
        coordinate_y(i,j)=j;
    end
end
[w z]=tforminv(tform,coordinate_x,coordinate_y);



%fundamental part!!!!->
for i=1:width
    for j=1:height
        source_x=w(i,j);
        source_y=z(i,j);
        if    (source_x>width-1    ||    source_y>height-1    ||
double(uint16(source_x))<=1 || double(uint16(source_y))<=1)
            hab(a(1,1),a(1,1))=hab(a(1,1),a(1,1))+1;
        else
            m=fix(source_x);
            n=fix(source_y);
            index_b=b(i,j);
            index_a0=a(m,n);
            index_a1=a(m+1,n);
            index_a2=a(m,n+1);
            index_a3=a(m+1,n+1);
            dx=source_x-m;
            dy=source_y-n;
            hab(index_a0,index_b)=hab(index_a0,index_b)+(1-dx)*(1-dy);
            hab(index_a1,index_b)=hab(index_a1,index_b)+dx*(1-dy);
            hab(index_a2,index_b)=hab(index_a2,index_b)+(1-dx)*dy;
            hab(index_a3,index_b)=hab(index_a3,index_b)+dx*dy;
        end
    end
```

```
end
habsum=sum(sum(hab));
index=find(hab~=0);
pab=hab/habsum;
Hab=sum(sum(-pab(index).*log2(pab(index))));
pa=sum(pab');
index=find(pa~=0);
Ha=sum(sum(-pa(index).*log2(pa(index))));
pb=sum(pab);
index=find(pb~=0);
Hb=sum(sum(-pb(index).*log2(pb(index))));
mi=Ha+Hb-Hab;
```

Here we got the result:

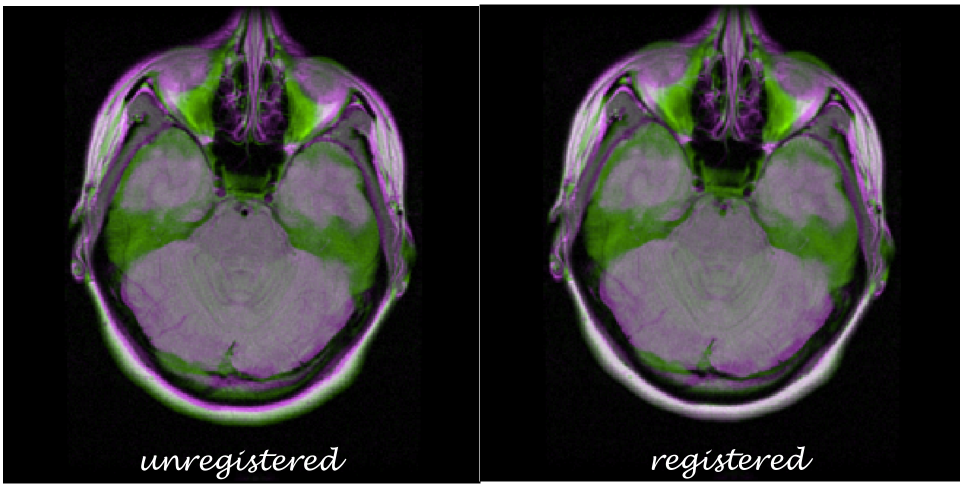


Fig 2

Ⅴ Evaluation of Ordinary MLE and JVHW estimator

In the process of MLE, there are several coefficiences in each iteration. However, these parameters are chosen randomly. Therefore, both the time complexity and the risks are multiplied due to the random selection. JVHW estimator, on the contrary, choose these coefficiences from delicate analysis of the samples and each time the iteration are more precious.
Fig 3 & 4 are the line charts of Mean Square Error of MLE and JVHW.
Table 3 & 4 are the forms of time consumption of each estimator.
From these statistics, we could figure out that:

[Tendency]

- As the size of sample capacity increased, the error of Mutual Information registration multiplies gradually.

[Comparison]

- Generally, the registration with JVHW estimator is far more precious than the normal MLE by more than 100 times.
- For each level of the size S of sample, the time consumption of JVHW and MLE is approximately the same.

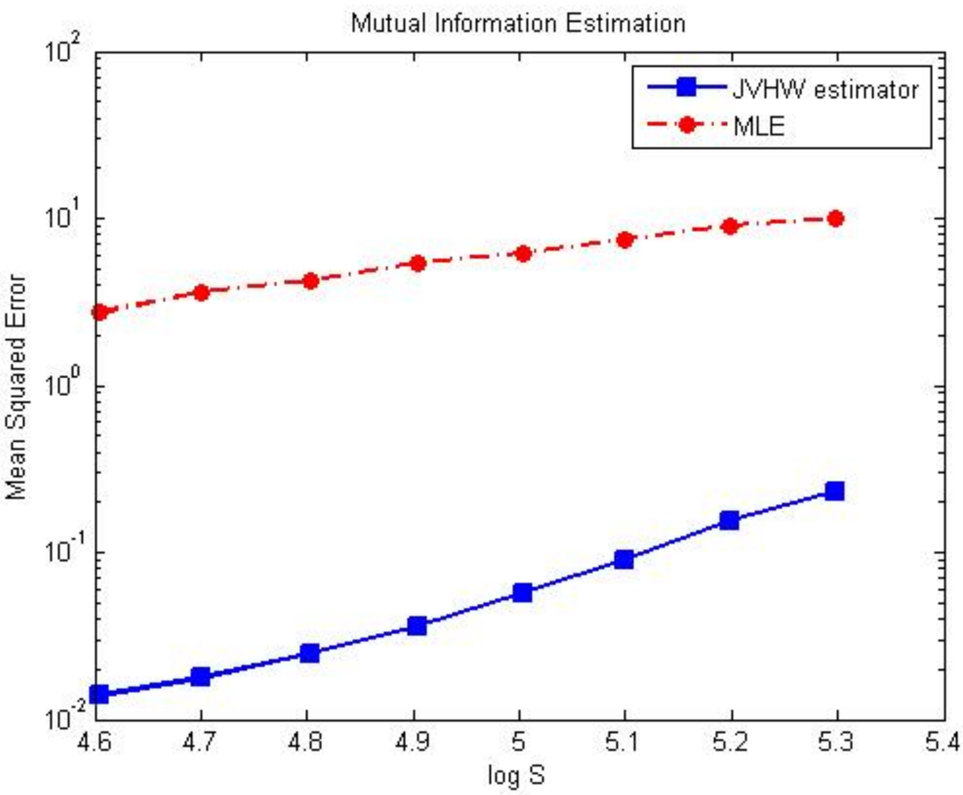


Fig 3

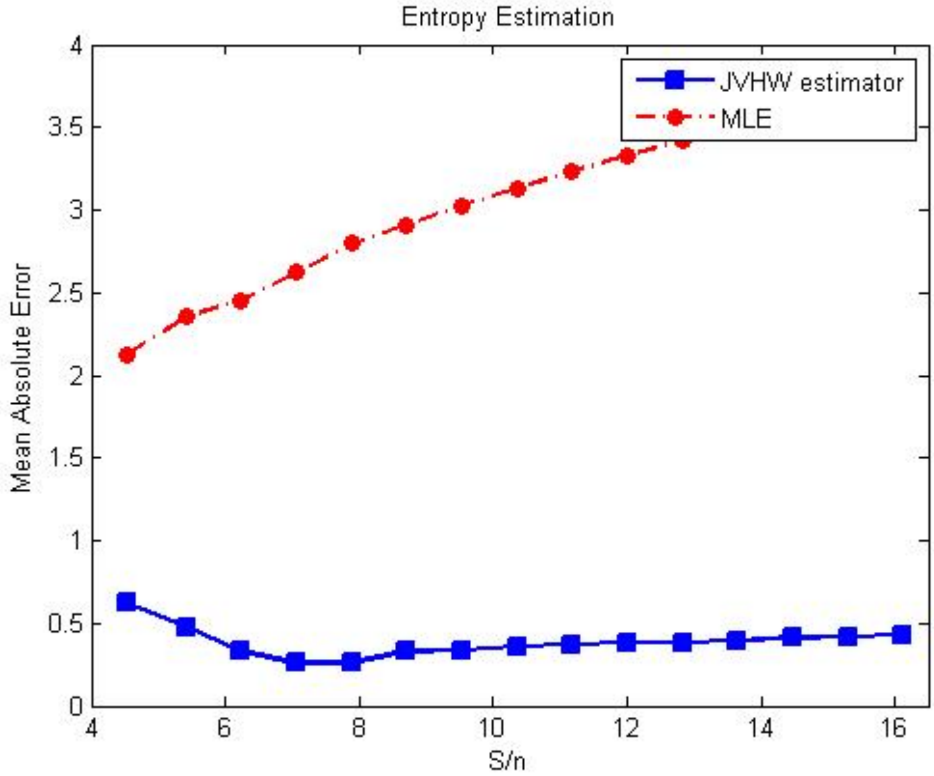


Fig 4

函数名称	调用	总时间	自用时间*	总时间图 (深色条带 = 自用时间)
test_MI	1	8.233 s	0.047 s	
est_MI_MLE	8	3.502 s	1.899 s	
est_MI_JVHW	8	3.421 s	1.817 s	
est_entro_JVHW	24	1.604 s	1.514 s	
est_entro_MLE	24	1.603 s	1.603 s	
randsmpl	8	0.562 s	0.192 s	

Table 3

函数名称	调用	总时间	自用时间*	总时间图 (深色条带 = 自用时间)
test_MI	1	6.000 s	0.020 s	
est_MI_JVHW	8	2.450 s	1.260 s	
est_MI_MLE	8	2.450 s	1.270 s	
est_entro_JVHW	24	1.190 s	1.090 s	
est_entro_MLE	24	1.180 s	1.180 s	
randsmpl	8	0.390 s	0.140 s	

Table 4

Ⅵ Summary

Virtually, the selection of algorithm and related operators or estimators are determined by the statistics. Elements including precision of the algorithm, time complexity, space complexity and size of the sample capacity should be taken into consideration. For the term of MIR, the precision of the registration is the most important standard in the evaluation of an algorithm. Therefore, JVHW estimator is more suitable rather than ordinary MLE.

Reference

[1] Ayman S.EI-Baz, Rajendra AcharyaU, Andrew F.Laine. Medical Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies [M]. ISBN 978-1-4419-8203-2. London: Springer Science+Business Media, 2011,

[2]J. Jiao, K. Venkat, Y. Han, and T. Weissman, "Minimax estimation of functions of discrete distributions," IEEE Transactions on Information Theory, vol. 61, no. 5, pp. 2835–2885, 2015. [3] T. A. Courtade and S. Verdú, "Cumulant generating function of codeword lengths in optimal lossless compression," in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jun./Jul. 2014, pp. 2494–2498.

[4] T. A. Courtade and S. Verdú, "Variable-length lossy compression and channel coding: Non-asymptotic converses via cumulant generating functions," in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jun./Jul. 2014, pp. 2499–2503.

[5] Mutual Information/Wikipedia https://en.wikipedia.org/wiki/Mutual_information

Working diary 4

I write my own program to compare the performance of the traditional mutual information based medical image registration methods, and the new ones that rely on JVHW entropy and mutual information estimators.

Discussion and Implement of Maximum Likelihood Estimator and JVHW Estimator: a report

Abstract: in this report stated the basic process of medical image registration in the environment of MATLAB with Maximum Likelihood Estimator and the JVHW estimator. Moreover, strengths and drawbacks together with comparison of these two methods are also discussed in detail. We want to illustrate that the JVHW estimators have advantages over Maximum Likelihood Estimator both in time complexity and precision.

Index Terms: Medical Image Registration, Maximum Likelihood Estimator, Mutual Information.

Ⅰ INTRODUCTION

Maximization of MI is a very general and powerful criterion, because no assumptions are made regarding the nature of this dependence and no limiting constraints are imposed on the image content of the modalities involved. The accuracy of the MI criterion is validated for rigid body registration of computed tomography (CT), magnetic resonance (MR), and photon emission tomography (PET) images by comparison with the stereotactic registration solution, while robustness is evaluated with respect to implementation issues, such as interpolation and optimization, and image content, including partial overlap and image degradation. [1]

To illustrate, Maximum Likelihood Estimator is a kind of output of Maximum Likelihood Estimation. Suppose in a random experiment with the results of A, B, C and D. Different conditions leads to distinct probability of the results in a single experiment. Therefore, in one specific random experiment with the result of A, we might assume that the probability of A is higher than any one of the others due to the inadequate times of experiments.

The probability of A is related to a parameter theta, θ. One of the main results of Maximum Likelihood Estimator is to give the functional equation of such parameters.

Maximum-likelihood estimation gives a unified approach to estimation, which is well-defined in the case of the normal distribution and many other problems. However, in some complicated problems,
difficulties do occur: in such problems, maximum-likelihood estimators are unsuitable or do not exist.

In the process of MLE, there are several coefficients in each iteration. However, these parameters are chosen randomly. Therefore, both the time complexity and the risks are multiplied due to the random selection. JVHW estimator, on the contrary, choose these coefficients from delicate analysis of the samples and each time the iteration are more precious.

Ⅱ THE PROCESS

Generally, entropy, more specifically Shannon entropy, refers to disorder or uncertainty. In information theory, systems are modeled by a transmitter, channel, and receiver. The transmitter produces messages that are sent through the channel. The channel modifies the message in some way. The receiver attempts to infer which message was sent. Entropy is the expected value (average) of the information contained in each message. "Messages" can be modeled by any flow of information.

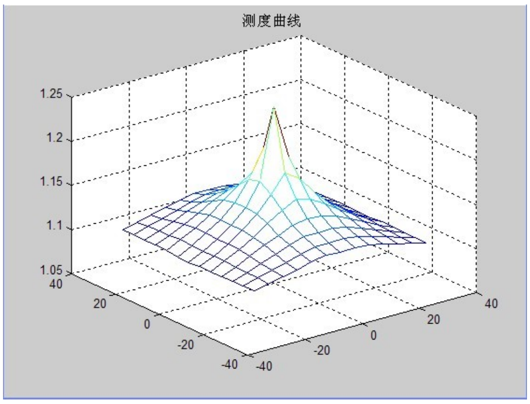
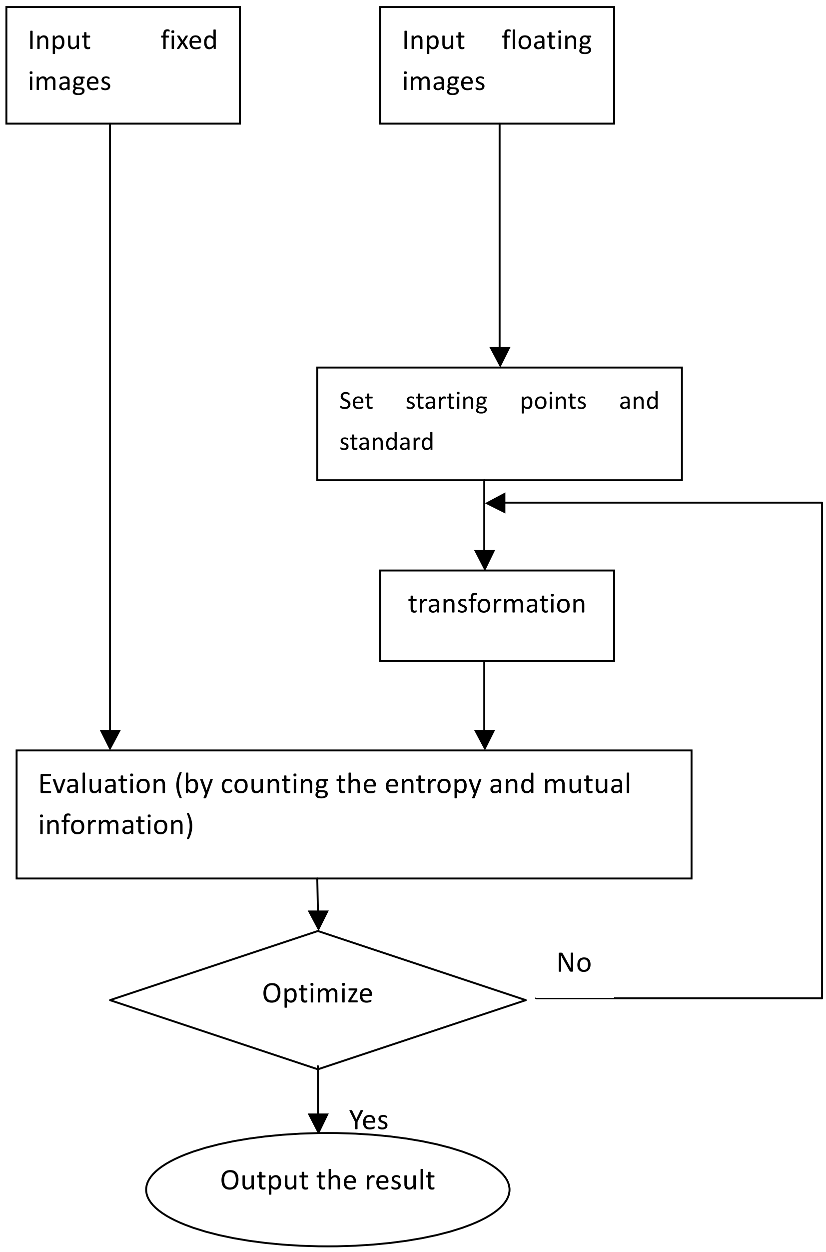


Fig1 the mutual information (entropy) which help to determine the precision of specific registration.

Ⅲ PROCESS AND ESTIMATOR



First we would give a brief structure of the whole process in Fig 2

In the procedure of the setting the standard, we use the Maximum Likelihood Estimation and JVHW estimator and discuss the difference of these two estimating

Here we attach a few of the most important parts of the two specific methods of the estimation:

```
f1nonzero = find(f(1,:) > 0);
lenf1nonzero = length(f1nonzero);
c_1 = zeros(1, wid);
if n > 15 && lenf1nonzero >0
    c_1(f1nonzero) = V * [ log(n) * ones(1,lenf1nonzero); log(f(1,f1nonzero)); ones(1,lenf1nonzero)];
    c_1 = max(c_1, 1/(1.9*log(n))); % make sure threshold is higher than 1/n
end
```

Here we estimate roughly the coefficients before we come into the precious registration.

2.1 Powell method

Powell's method, strictly Powell's conjugate direction method, is an algorithm proposed by Michael J. D. Powell for finding a local minimum of a function. The function need not be differentiable, and no derivatives are taken.

The function must be a real-valued function of a fixed number of real-valued inputs. The caller passes in the initial point. The caller also passes in a set of initial search vectors. Typically N search vectors are passed in which are simply the normals aligned to each axis.

The method minimises the function by a bi-directional search along each search vector, in turn. The new position can then be expressed as a linear combination of the search vectors. The new displacement vector becomes a new search vector, and is added to the end of the search vector list. Meanwhile, the search vector which contributed most to the new direction, i.e. the one which was most successful, is deleted from the search vector list. The algorithm iterates an arbitrary number of times until no significant improvement is made.

The method is useful for calculating the local minimum of a continuous but complex function, especially one without an underlying mathematical definition, because it is not necessary to take derivatives. The basic algorithm is simple; the complexity is in the linear searches along the search vectors, which can be achieved via Brent's method.

Here we use this algorithm to connect the estimation together with the registration process:

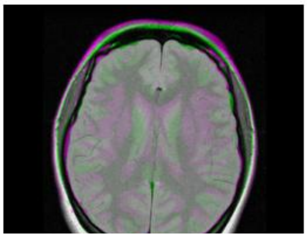
```
function [mi]=PV(x,y,ang,handles)
a=handles.data;
a=double(a);
```

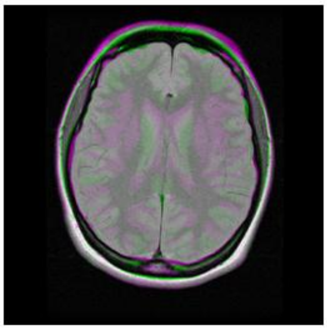
```
b=handles.data2;
b=double(b);
[M,N]=size(a);
hab=zeros(256,256);
ha=zeros(1,256);
hb=zeros(1,256);
if max(max(a))~=min(min(a))
    a=(a-min(min(a)))/(max(max(a))-min(min(a)));
else
    a=zeros(M,N);
end
if max(max(b))~=min(min(b))
    b=(b-min(min(b)))/(max(max(b))-min(min(b)));
else
    b=zeros(M,N);
end
a=double(int16(a*255))+1;
b=double(int16(b*255))+1;
[width,height]=size(b);
u=(width-1)/2;
v=(height-1)/2;
rad=pi/180*ang;
t1=[1 0 0;0 1 0;x y 1];
t2=[1 0 0;0 1 0;-u -v 1];
t3=[cos(rad) -sin(rad) 0;sin(rad) cos(rad) 0;0 0 1];
t4=[1 0 0;0 1 0;u v 1];
T=t2*t3*t4*t1;
tform=maketform('affine',T);
coordinate_x=zeros(width,height);
coordinate_y=zeros(width,height);
for i=1:width
    for j=1:height
        coordinate_x(i,j)=i;
    end
end
for i=1:width
    for j=1:height
        coordinate_y(i,j)=j;
    end
end
[w z]=tforminv(tform,coordinate_x,coordinate_y);
for i=1:width
    for j=1:height
        source_x=w(i,j);
```

```
            source_y=z(i,j);
            if(source_x>width-1||source_y>height-1||double(uint16(source_x))<=1||double(uint16(source_y))<=1)
                hab(a(1,1),a(1,1))=hab(a(1,1),a(1,1))+1;
            else
                m=fix(source_x);
                n=fix(source_y);
                index_b=b(i,j);
                index_a0=a(m,n);
                index_a1=a(m+1,n);
                index_a2=a(m,n+1);
                index_a3=a(m+1,n+1);
                dx=source_x-m;
                dy=source_y-n;
                hab(index_a0,index_b)=hab(index_a0,index_b)+(1-dx)*(1-dy);
                hab(index_a1,index_b)=hab(index_a1,index_b)+dx*(1-dy);
                hab(index_a2,index_b)=hab(index_a2,index_b)+(1-dx)*dy;
                hab(index_a3,index_b)=hab(index_a3,index_b)+dx*dy;
            end
        end
    end
    habsum=sum(sum(hab));
    index=find(hab~=0);
    pab=hab/habsum;
    Hab=sum(sum(-pab(index).*log2(pab(index))));
    pa=sum(pab,2);
    index=find(pa~=0);
    Ha=sum(sum(-pa(index).*log2(pa(index))));
    pb=sum(pab,1);
    index=find(pb~=0);
    Hb=sum(sum(-pb(index).*log2(pb(index))));
    mi=Ha+Hb-Hab;
```

Ⅲ RESULTS

The following is the results of the process, here we select two brain images as the input:





Time consumption MLE: 0.020s
Time consumption JVHW: 0.027s

Result of the MLE:

Result of the JVHW:

Here we could hardly figure out the difference of the difference between the two estimation by simply the observation.

However, we can use the Mean Square Error to evaluate the two images.

Followed is the source code:

```
for iter = num:-1:1
      S = record_S(iter);
      px = betarnd(0.6,0.5,S,1);
      px = px/sum(px);
      pz = betarnd(0.6,0.5,S,1);
      pz = pz/sum(pz);
      py_cond_x = bsxfun(@circshift, pz, 0:S-1).';
      pxy = diag(px)*py_cond_x;
      [X,Y] = ind2sub([S,S],randsmpl(pxy(:),n, mc_times));
      true_S(iter) = MI_true(pxy);
  record_JVHW = est_MI_JVHW(X,Y);
  record_MLE = est_MI_MLE(X,Y);
      JVHW_S(iter) = mean((record_JVHW - true_S(iter)).^2);
      MLE_S(iter) = mean((record_MLE - true_S(iter)).^2);
End
```

From the result:
Mean Square Error MLE: 1.023×10^{-8}
Mean Square Error JVHW: 3.123×10^{-12}

Ⅳ SUMMARY

Here we could figure out that:
1 the time consumption of MLE is a little smaller than JVHW.
2 the prevision of MLE is far more larger than JVHW.

Reference

[1] Ayman S.EI-Baz, Rajendra AcharyaU, Andrew F.Laine. Medical Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies [M]. ISBN 978-1-4419-8203-2. London: Springer Science+Business Media, 2011,
[2]J. Jiao, K. Venkat, Y. Han, and T. Weissman, “Minimax estimation of functions of discrete distributions,” IEEE Transactions on Information Theory, vol. 61, no. 5, pp. 2835–2885, 2015. [3] T. A. Courtade and S. Verdú, “Cumulant generating function of codeword lengths in optimal lossless compression,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jun./Jul. 2014, pp. 2494–2498.
[4] T. A. Courtade and S. Verdú, “Variable-length lossy compression and channel coding: Non-asymptotic converses via cumulant generating functions,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jun./Jul. 2014, pp. 2499–2503.
[5] Mutual Information/Wikipedia https://en.wikipedia.org/wiki/Mutual_information
[6]Gonzalez. Digital Image Processing (Second Edition)[M]. ISBN 978-3-642-15815-5. Berlin: Springer-Verlag Berlin Heidelberg, 2011.
[7]MATLAB functions descriptions