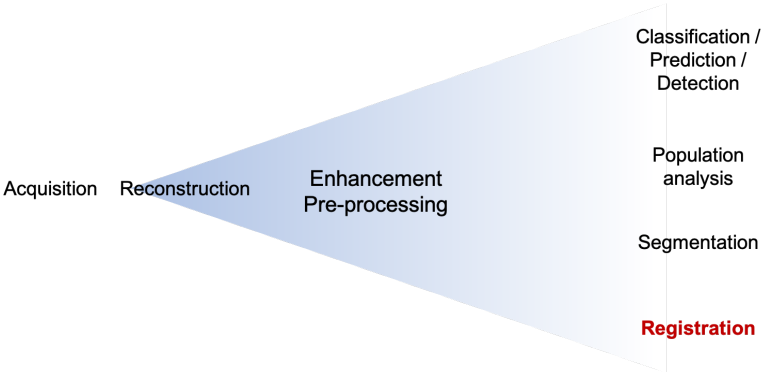
Loss metrics and registration algorithms

Ender Konukoglu

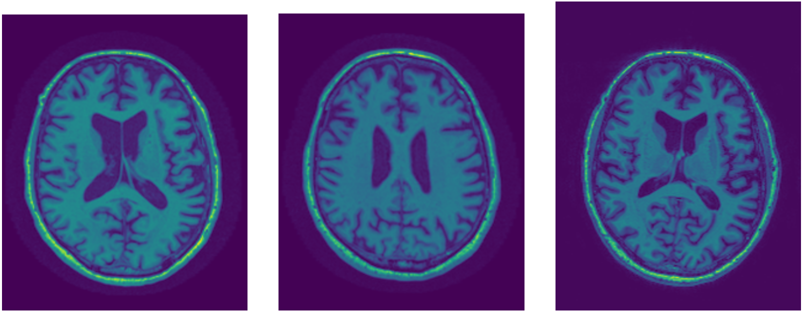
ETH Zürich

March 17, 2020

Registration is an essential task in medical image analysis



What is image registration and why do we need it?



Time point 1 Time point 2 Time point 3

A common problem

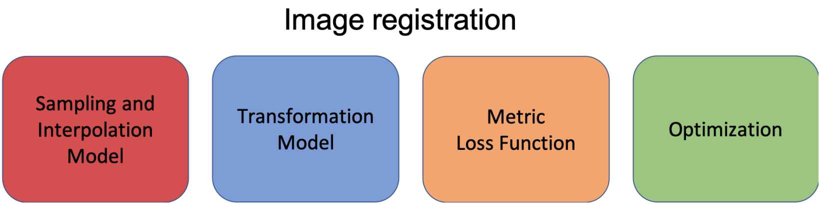
These are three images taken from the same individual 6 months apart. Images show the same cross-section (90th slide) from three different volumetric images. Notice that they do not show the same anatomy. That is because they are not aligned. Image registration aligns these images through spatial transformations and allows comparisons.

Useful for many different applications

Image Registration
Spatial normalization

- Aligning images of different modalities
- Longitudinal analysis
- Atlas-based segmentation
- Population Analysis
- Image analysis for interventions: alignment of pre- and intra-intervention images
- Any other analysis that requires aligning different images

Four main components



Any registration algorithm is composed of four components. We have seen:

- Sampling and interpolation model
- Transformation models

Today we will study

- Metric / Loss functions
- Optimization - briefly

Outline

- Classifying registration problems
- Landmark-based registration
  - Landmarks
  - Linear registration
  - Thin Plate Splines
- Intensity-based registration
  - Intra-modality loss metrics
  - Inter-modality loss metrics
- Optimization - briefly
  - Regularization
  - Gradient descent
  - Sequential optimization

Section 2

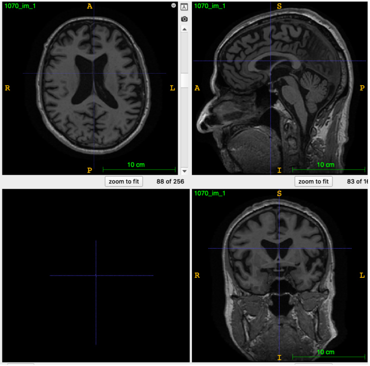
Classifying registration problems

Classification in two axes

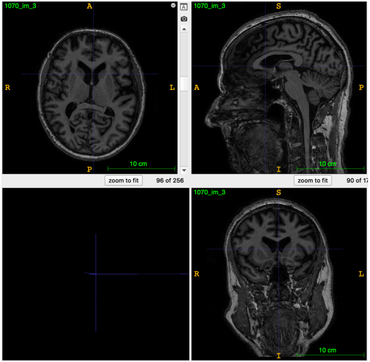
	Same subject's images	Different subjects' images
Same modality	Intra-subject, intra-modality	Inter-subject, intra-modality
Different modality	Intra-subject, inter-modality	Inter-subject, inter-modality

- Registering images of the same or different subjects
  - Alignment strategy would change
  - Transformations are different
- Registering similar or different modalities
  - Need to use different loss functions
  - Intensity-based loss functions that are not affected by differences in modality variations
- Colors indicate the difficulty of the registration problem.
  - Green is the easiest and red is the hardest.
  - Yellow is easier than the orange but this is a subjective judgement
  - Difficulty also depends on the anatomy, e.g. brain is easier than abdomen

Intra-subject registration



Time point #1

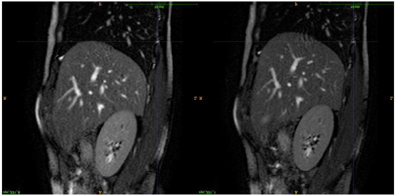


Time point #2

Aligning two different images of the same person. Many application areas in research and in practice.

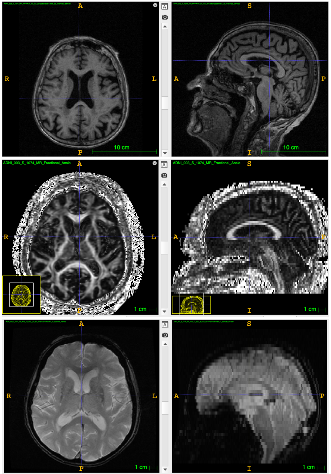
Easier problem than inter-subject because we expect the same anatomy.

Intra-subject registration can be difficult



3D+time lung MRI, cross-sections at two time points.

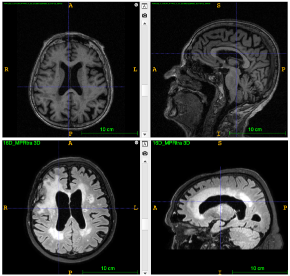
- The deformation of the anatomy during motion can be very complex [see lung MRI video in Moodle]. Transformation models may not be able to model this accurately. An example for this is the breathing motion and its effect on the abdomen.
- Images of the same individual may differ vastly in intensity characteristics. Neuroimaging studies are a good example of this.



Three different MRI contrasts of the same individual. Top to bottom: Structural, Fractional Anisotropy, T2star.

Inter-subject registration is often difficult

CT images of different individuals showing the same FOV.

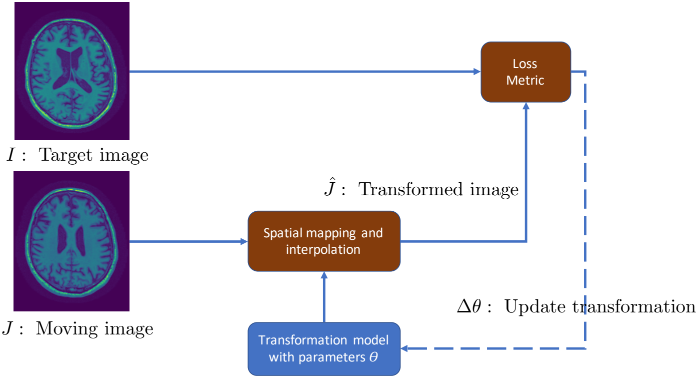


Top: image from a healthy subject, Bottom: image from a patient with a brain tumor.

- Differences in anatomies even when images show the same structures.
- **Lack of exact correspondence between two images.**
- Presence of pathologies make registration even more difficult.

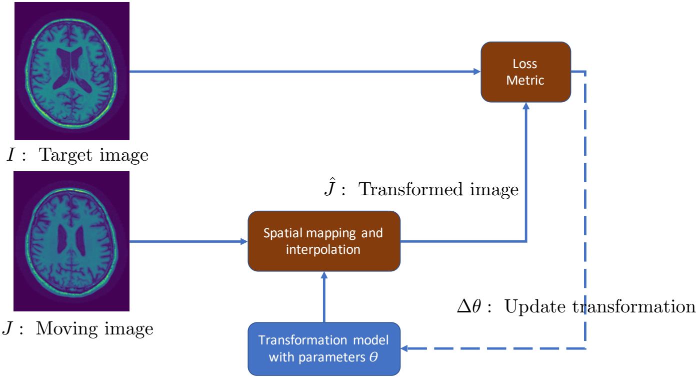
Outline

Overview on a registration algorithm



$$\theta^* = \arg_\theta \min \mathcal{L}(\theta)$$

Overview on a registration algorithm



$$\theta^* = \arg_\theta \min \mathcal{L}(\theta)$$

Main questions of today

What loss metrics $\mathcal{L}(\theta)$ do we use and how do we optimize, i.e. solve the $\arg_\theta \min$ problem?

Section 3

Landmark-based registration

Landmarks

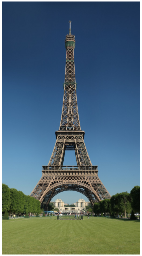
Landmark [from Oxford English Dictionary `www.oed.com`]

An object in the landscape, which, by its conspicuousness, serves as a guide in the direction of one's course (originally and esp. as a guide to sailors in navigation); hence, any conspicuous object which characterizes a neighbourhood or district.

Landmarks

Landmark [from Oxford English Dictionary `www.oed.com`]

An object in the landscape, which, by its conspicuousness, serves as a guide in the direction of one's course (originally and esp. as a guide to sailors in navigation); hence, any conspicuous object which characterizes a neighbourhood or district.

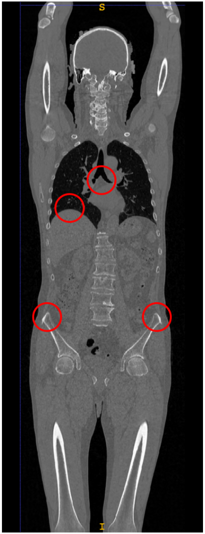


Landmarks in medical images

Landmarks in medical imaging

- prominent locations in an image
- distinct from their surroundings, similar to interest points from computer vision
- easily identifiable
- can be used to orient ourselves in an image
- **landmarks will be used to define correspondence between images**
- landmark definition depends on the imaged structure and type of images to align
- most commonly manually placed

Landmarks in medical images



Landmarks in medical imaging

- prominent locations in an image
- distinct from their surroundings, similar to interest points from computer vision
- easily identifiable
- can be used to orient ourselves in an image
- **landmarks will be used to define correspondence between images**
- landmark definition depends on the imaged structure and type of images to align
- most commonly manually placed

Examples for thorax

- Trachea bifurcation
- Most superior tip of the liver
- Tip of the pelvis bones - iliac crests

Landmarks should be identifiable across images for defining correspondence

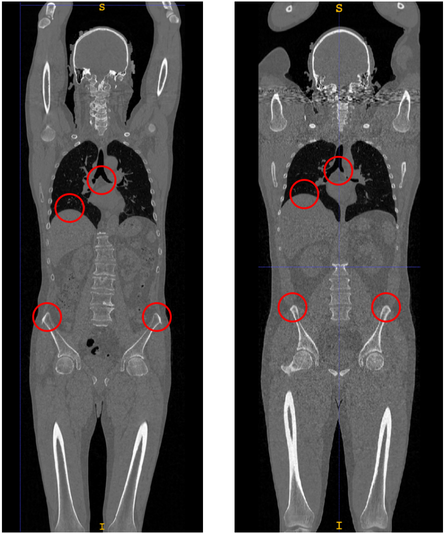
Subject #1

Subject #2

Landmarks in medical imaging

- should be identifiable across images
- same landmark in both of the images
- corresponding landmaks will guide registration procedure and help align the images
- correspondence between landmarks define correspondence between different pixels / voxels

Landmarks should be identifiable across images for defining correspondence

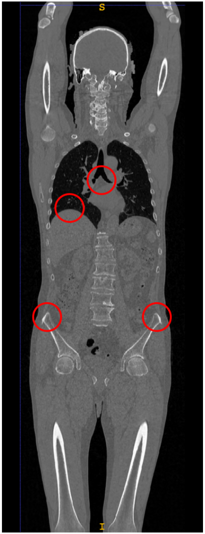


Subject #1 Subject #2

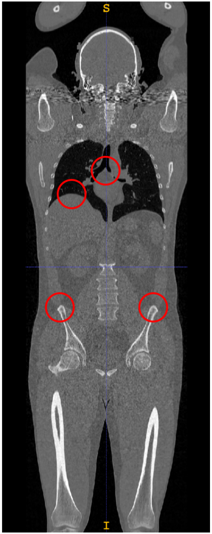
Landmarks in medical imaging

- should be identifiable across images
- same landmark in both of the images
- corresponding landmaks will guide registration procedure and help align the images
- correspondence between landmarks define correspondence between different pixels / voxels

Landmarks should be identifiable across images for defining correspondence



Subject #1

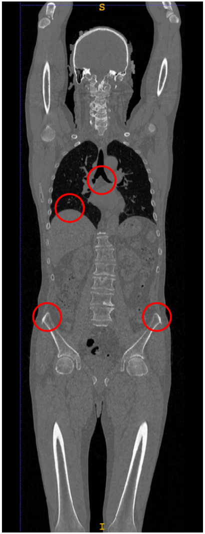


Subject #2

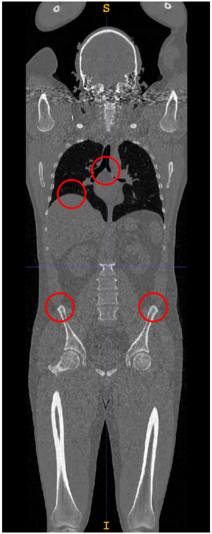
Landmarks in medical imaging

- should be identifiable across images
- same landmark in both of the images
- corresponding landmaks will guide registration procedure and help align the images
- correspondence between landmarks define correspondence between different pixels / voxels
- across which images?

Landmarks should be identifiable across images for defining correspondence



Subject #1

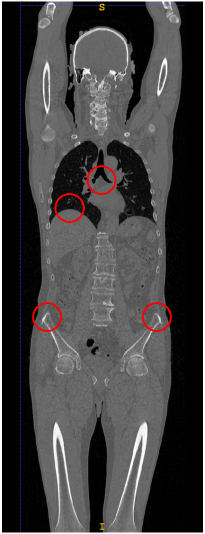


Subject #2

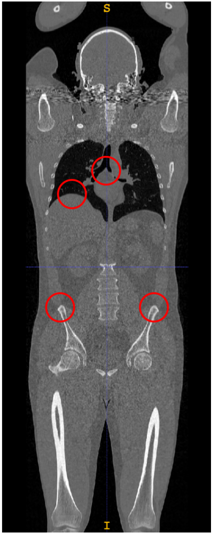
Landmarks in medical imaging

- should be identifiable across images
- same landmark in both of the images
- corresponding landmaks will guide registration procedure and help align the images
- correspondence between landmarks define correspondence between different pixels / voxels
- across which images?
- that depends on the application

Same modality, different subjects



Subject #1



Subject #2

- both CT images acquired with similar protocols, e.g. without contrast
- images of different people
- natural landmark positions are anatomical structures common to every person
- more generally common to every sample

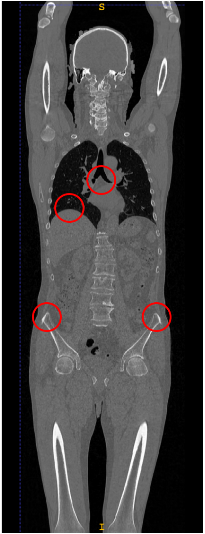
Different modality, different subjects

Subject #1

Subject #2

- different modalities: one CT and one MRI
- images of different people
- showing the same anatomical area
- natural landmark positions are anatomical structures common to every sample

Different modality, different subjects

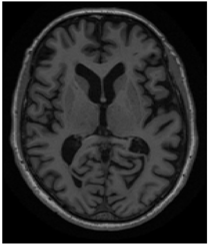


Subject #1

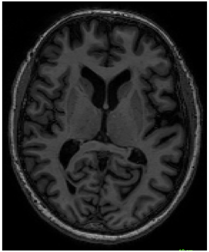
Subject #2

- different modalities: one CT and one MRI
- images of different people
- showing the same anatomical area
- natural landmark positions are anatomical structures common to every sample
- landmarks should be identifiable in both of the modalities

Same person, same modality



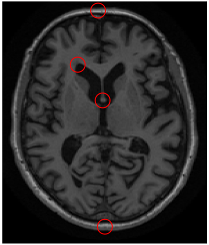
Time point #1



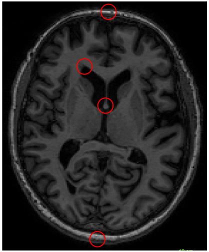
Time point #2

- same modality - MRI
- images of the same person taken at different time points

Same person, same modality



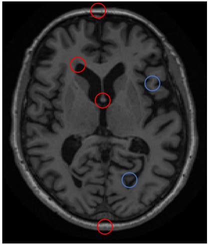
Time point #1



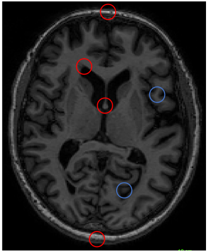
Time point #2

- same modality - MRI
- images of the same person taken at different time points
- usual landmark locations - anatomical structures common to everyone

Same person, same modality



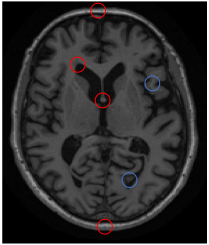
Time point #1



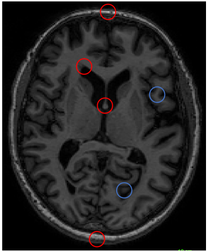
Time point #2

- same modality - MRI
- images of the same person taken at different time points
- usual landmark locations - anatomical structures common to everyone
- in addition, sample-specific structural variations - see blue points
- sample-specific landmarks will lead to more accurate allignment

Same person, same modality



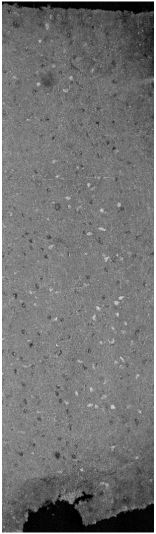
Time point #1



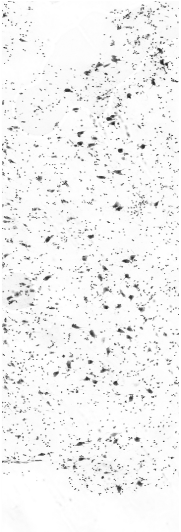
Time point #2

- same modality - MRI
- images of the same person taken at different time points
- usual landmark locations - anatomical structures common to everyone
- in addition, sample-specific structural variations - see blue points
- sample-specific landmarks will lead to more accurate allignment
- same person and different modalities is also very common and same idea applies

Landmarks in biological samples



OCT

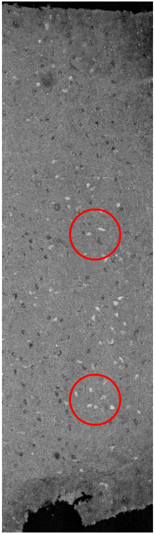


Nissl

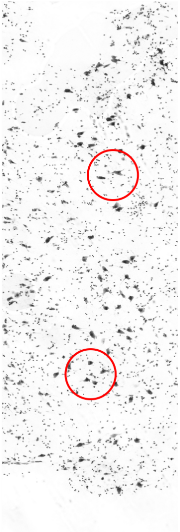
Even though we talked about medical images so far, the same concepts apply to biological images as well.

- different modalities - Optical Coherence Tomography and Nissl-stained microscopy image
- image of the same tissue sample

Landmarks in biological samples



OCT

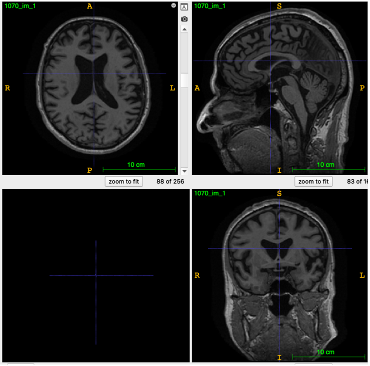


Nissl

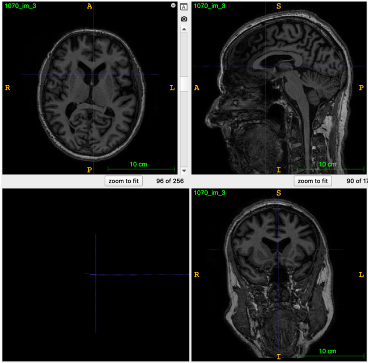
Even though we talked about medical images so far, the same concepts apply to biological images as well.

- different modalities - Optical Coherence Tomography and Nissl-stained microscopy image
- image of the same tissue sample
- corresponding structures
- constellations of neurons in these examples is used
- defining such landmarks across different tissue samples can be difficult

Landmarks in 3D



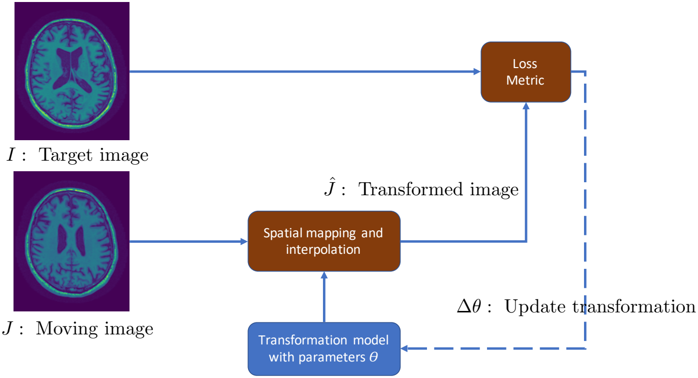
Time point #1



Time point #2

Although we discussed landmarks in 2D, medical images are mostly in 3D. Therefore, landmarks need to be defined in 3D.

How to align images with the landmarks?

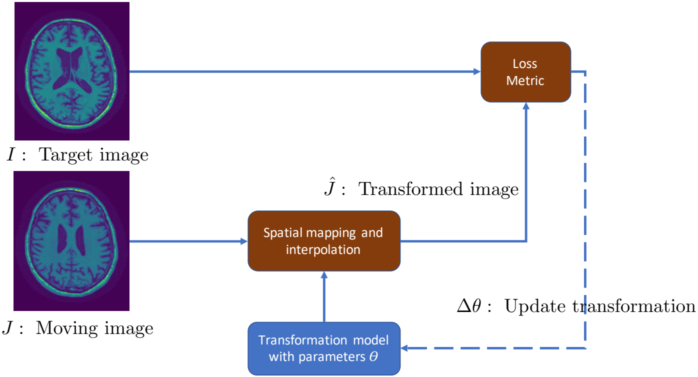


$$\theta^* = \arg_\theta \min \mathcal{L}(\theta)$$

Main questions of today

What loss metrics $\mathcal{L}(\theta)$ do we use and how do we optimize, i.e. solve the $\arg_\theta \min$ problem?

How to align images with the landmarks?

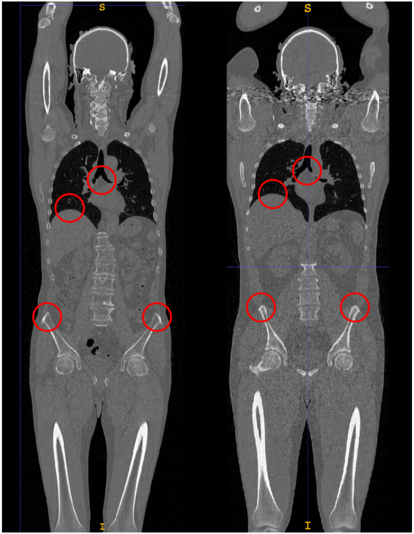


$$\theta^* = \arg_\theta \min \mathcal{L}(\theta)$$

Landmark-based loss function

$$\mathcal{L}(\theta) = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta \left(\mathbf{x}_n^{(2)} \right) \right\|^2, \; \mathbf{x} \in \mathbb{R}^d$$

Breaking down the cost function



$$\mathcal{L} = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta\left(\mathbf{x}_n^{(2)}\right) \right\|_2^2, \quad \mathbf{x} \in \mathbb{R}^d$$

- $\mathbf{x}_n^{(1)}$ is the nth landmark in the first image
- $\mathbf{x}_n^{(2)}$ is the corresponding landmark in the second image
- $T_\theta\left(\mathbf{x}_n^{(2)}\right)$ is the transformed version of $\mathbf{x}_n^{(2)}$
- θ is the set of transformation parameters
- d is the dimension of the problem, e.g. 2D, 3D, 3D+time, ...
- $\| \cdot \|_2^2$ s the usual ℓ_2 loss
- other types of losses are also possible

Subsection 1

Landmark-based registration - Linear alignment

Aligning with linear transformations

$$\mathcal{L}(\theta) = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta\left(\mathbf{x}_n^{(2)}\right) \right\|_2^2, \ \mathbf{x} \in \mathbb{R}^d$$

When aligning with linear transformation T_θ is defined through a matrix $\mathbf{A}$ and a translation $\mathbf{t}$

Aligning with linear transformations

$$\mathcal{L}(\theta) = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta\left(\mathbf{x}_n^{(2)}\right) \right\|_2^2, \ \mathbf{x} \in \mathbb{R}^d$$

When aligning with linear transformation T_θ is defined through a matrix $\mathbf{A}$ and a translation $\mathbf{t}$

$$\mathcal{L}(\mathbf{A}, \mathbf{t}) = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - \left(\mathbf{A}\mathbf{x}_n^{(2)} + \mathbf{t}\right) \right\|_2^2, \ \mathbf{x} \in \mathbb{R}^d$$

Aligning with linear transformations

$$\mathcal{L}(\theta) = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta\left(\mathbf{x}_n^{(2)}\right) \right\|_2^2, \; \mathbf{x} \in \mathbb{R}^d$$

When aligning with linear transformation T_θ is defined through a matrix $\mathbf{A}$ and a translation $\mathbf{t}$

$$\mathcal{L}(\mathbf{A}, \mathbf{t}) = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - \left(\mathbf{A}\mathbf{x}_n^{(2)} + \mathbf{t}\right) \right\|_2^2, \; \mathbf{x} \in \mathbb{R}^d$$

The loss is minimized to determine the best transformation parameters $\mathbf{A}$ and $\mathbf{t}$

$$\arg_{\mathbf{A},\mathbf{t}} \min \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - \left(\mathbf{A}\mathbf{x}_n^{(2)} + \mathbf{t}\right) \right\|_2^2$$

Aligning with linear transformations

$$\mathcal{L}(\theta) = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta\left(\mathbf{x}_n^{(2)}\right) \right\|_2^2, \ \mathbf{x} \in \mathbb{R}^d$$

When aligning with linear transformation T_θ is defined through a matrix $\mathbf{A}$ and a translation $\mathbf{t}$

$$\mathcal{L}(\mathbf{A}, \mathbf{t}) = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - \left(\mathbf{A}\mathbf{x}_n^{(2)} + \mathbf{t}\right) \right\|_2^2, \ \mathbf{x} \in \mathbb{R}^d$$

The loss is minimized to determine the best transformation parameters $\mathbf{A}$ and $\mathbf{t}$

$$\arg_{\mathbf{A},\mathbf{t}} \min \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - \left(\mathbf{A}\mathbf{x}_n^{(2)} + \mathbf{t}\right) \right\|_2^2$$

Exercise

The above minimization's solution can be determined in an analytical form. Can you derive the result of the optimization?

Aligning with linear transformations

$$\mathcal{L}(\theta) = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta \left(\mathbf{x}_n^{(2)} \right) \right\|_2^2, \ \mathbf{x} \in \mathbb{R}^d$$

When aligning with linear transformation T_θ is defined through a matrix **A** and a translation **t**

$$\mathcal{L}(\mathbf{A}, \mathbf{t}) = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - \left(\mathbf{A}\mathbf{x}_n^{(2)} + \mathbf{t} \right) \right\|_2^2, \ \mathbf{x} \in \mathbb{R}^d$$

The loss is minimized to determine the best transformation parameters **A** and **t**

$$\arg_{\mathbf{A},\mathbf{t}} \min \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - \left(\mathbf{A}\mathbf{x}_n^{(2)} + \mathbf{t} \right) \right\|_2^2$$

Exercise

The above minimization's solution can be determined in an analytical form. Can you derive the result of the optimization?

For other loss functions the minimization can be solved using optimization algorithms.

Example I - Rigid transformation, same subject, same modality

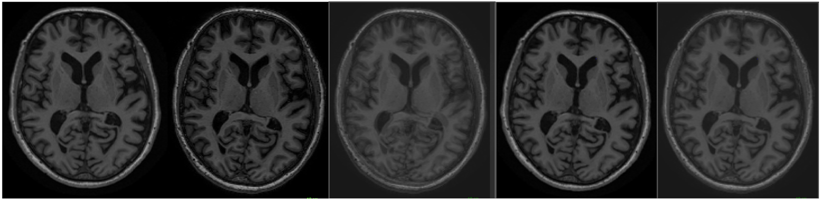


Figure: From left to right: time point #1, time point #2, overlay, time point #1 registered with rigid registration with landmarks, overlay after registration

- Results from rigid registration - T_θ is a rigid transformation, i.e. rotation and translation.
- Using only 4 landmarks the rigid registration correctly aligns the images
- Notice that the overlay is sharper - meaning good alignment

Example II - Rigid transformation, different subject, same modality

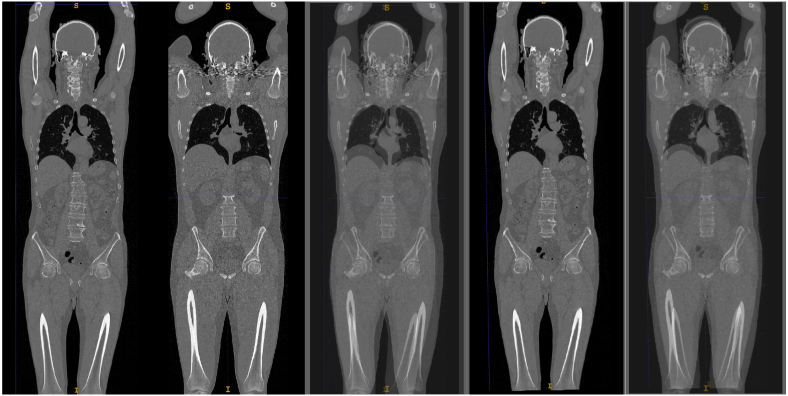


Figure: From left to right: subject #1, subject #2, overlay, subject #1 registered with rigid registration with landmarks, overlay after registration

- Rigid registration may not be enough to account for differences between different subjects.
- Landmarks: trachea bifurcation, liver tip and pelvis tips
- Notice the landmarks do not match perfectly.

Example III - Affine transformation, different subject, same modality

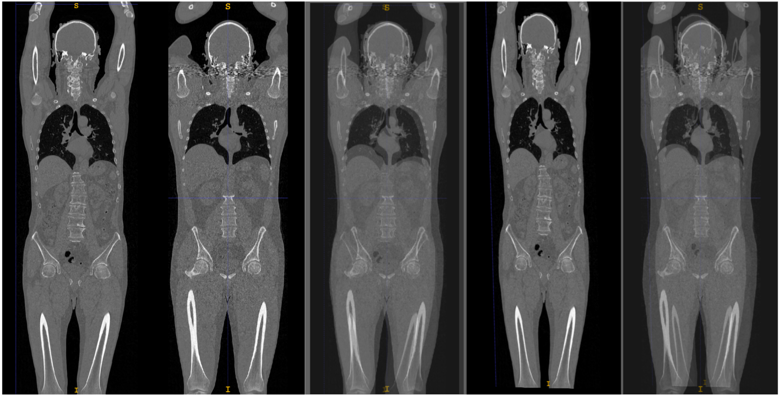
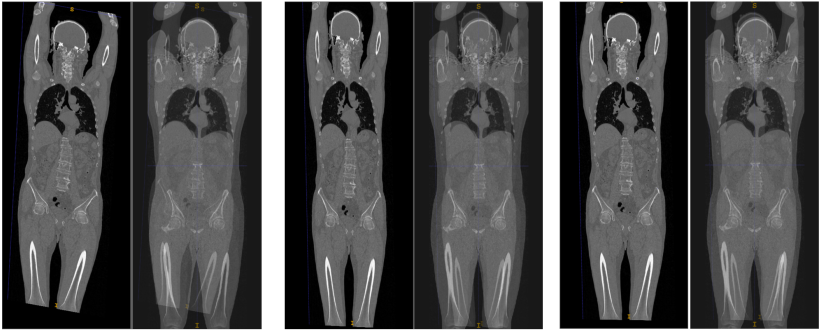


Figure: From left to right: subject #1, subject #2, overlay, subject #1 registered with affine registration with landmarks, overlay after registration

- Affine registration may be a bit better in accounting for inter-subject differences.
- Landmarks: trachea bifurcation, liver tip and pelvis tips
- Notice the landmarks match better.

Analysis: How would adding more landmarks change?

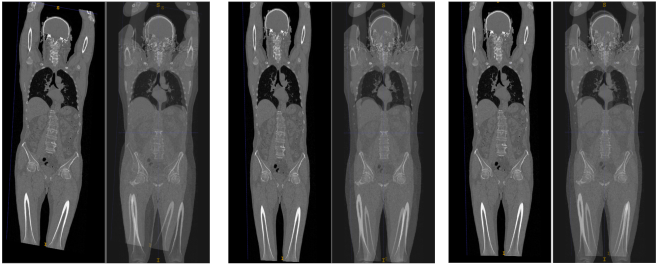


Three landmarks Four landmarks Seven landmarks

Figure: In each pair, the moving image and the overlay after alignment is shown. The first three landmarks are the tip of liver, left pelvis tip and the trachea bifurcation.

- At three landmarks, alignment for the landmarks are perfect but for the other parts it is not.
- That is because affine transformation has 6 unknowns, and 6 equalities - given by three landmarks - gives an well-determined system of equations.
- Increasing number of landmarks yield worst alignment for the landmarks but better for other areas.
- More landmarks leads to an over-determined system of equations.

Constrained transformation

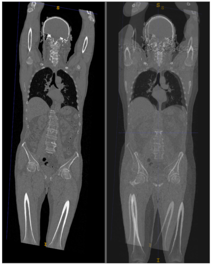


Three landmarks | Four landmarks | Seven landmarks

Figure: In each pair, the moving image and the overlay after alignment is shown. The first three landmarks are the tip of liver, left pelvis tip and the trachea bifurcation.

Why does the affine transformation not align all the landmarks perfectly?

Constrained transformation



Three landmarks

Four landmarks

Seven landmarks

Figure: In each pair, the moving image and the overlay after alignment is shown. The first three landmarks are the tip of liver, left pelvis tip and the trachea bifurcation.

Why does the affine transformation not align all the landmarks perfectly?
With 6 parameters, it does not have the degrees of freedom to satisfy more than 3 correspondences in 2D. Note that 3 correspondences lead to 6 equations; each coordinate will give one equation.

$$\mathcal{L} = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - \left(\mathbf{A}\mathbf{x}_n^{(2)} + \mathbf{t} \right) \right\|_2^2, \ \mathbf{x} \in \mathbb{R}^d$$

Subsection 2

Landmark-based registration - Non-linear alignment with Thin Plate Splines

Reminder - Thin plate splines (TPS)

Radial basis functions model:

$$u(x') = \sum_{n=1}^{N} \phi(|x' - x'_n|) d_n, \ \phi : \mathbb{R}^+ \to \mathbb{R}, \ d_n \in \mathbb{R}^{2 \text{ or } 3}$$

where x_n are the control points, ϕ are the basis functions and d_n are the non-linear coefficients.

In TPS, the basis functions are defined as

$$\phi(|x' - x'_n|) = \phi(r) = r^2 \log r$$

- Sometimes defined together with an affine transformation as follows

  $$x = \mathbf{A}x' + t + \sum_{n=1}^{N} \phi(|x' - x'_n|) d_n,$$

  where $\mathbf{A}$ and t define a global linear transformation and TPS define local refinement.
- Control points do not have to be uniformly spaced, they can be randomly dispersed.
- TPS is extensively used, especially for landmark-based registration.

Aligning with non-linear transformations

$$\mathcal{L} = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta \left(\mathbf{x}_n^{(2)} \right) \right\|_2^2, \ \mathbf{x} \in \mathbb{R}^d$$

When aligning with TPS, the cost function becomes

Aligning with non-linear transformations

$$\mathcal{L} = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta \left(\mathbf{x}_n^{(2)} \right) \right\|_2^2, \ \mathbf{x} \in \mathbb{R}^d$$

When aligning with TPS, the cost function becomes

$$\mathcal{L} = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - \mathbf{x}_n^{(2)} - \sum_{k=1}^{N} \phi \left(\mathbf{x}_n^{(2)} - \mathbf{x}_k^{(2)} \right) d_k \right\|_2^2, \ \mathbf{x} \in \mathbb{R}^d$$

Note that the displacement vectors are defined through the same landmarks.

Aligning with non-linear transformations

$$\mathcal{L} = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta\left(\mathbf{x}_n^{(2)}\right) \right\|_2^2, \;\; \mathbf{x} \in \mathbb{R}^d$$

When aligning with TPS, the cost function becomes

$$\mathcal{L} = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - \mathbf{x}_n^{(2)} - \sum_{k=1}^{N} \phi\left(\mathbf{x}_n^{(2)} - \mathbf{x}_k^{(2)}\right) d_k \right\|_2^2, \;\; \mathbf{x} \in \mathbb{R}^d$$

Note that the displacement vectors are defined through the same landmarks. Similar to linear, the loss is minimized to determine the best transformation parameters **A** and **t**

$$\arg_{\{d_k\}_{k=1,\ldots,N}} \min \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - \mathbf{x}_n^{(2)} - \sum_{k=1}^{N} \phi\left(\mathbf{x}_n^{(2)} - \mathbf{x}_k^{(2)}\right) d_k \right\|_2^2$$

Aligning with non-linear transformations

$$\mathcal{L} = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta\left(\mathbf{x}_n^{(2)}\right) \right\|_2^2, \ \mathbf{x} \in \mathbb{R}^d$$

When aligning with TPS, the cost function becomes

$$\mathcal{L} = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - \mathbf{x}_n^{(2)} - \sum_{k=1}^{N} \phi\left(\mathbf{x}_n^{(2)} - \mathbf{x}_k^{(2)}\right) d_k \right\|_2^2, \ \mathbf{x} \in \mathbb{R}^d$$

Note that the displacement vectors are defined through the same landmarks. Similar to linear, the loss is minimized to determine the best transformation parameters **A** and **t**

$$\arg_{\{d_k\}_{k=1,\ldots,N}} \min \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - \mathbf{x}_n^{(2)} - \sum_{k=1}^{N} \phi\left(\mathbf{x}_n^{(2)} - \mathbf{x}_k^{(2)}\right) d_k \right\|_2^2$$

After determining the optimal parameters d_k^* one can use it to find the displacement vectors everywhere in the target image, not just the landmarks

$$T_\theta(\mathbf{x}^{(2)}) = \mathbf{x}^{(2)} + \sum_{k=1}^{N} \phi\left(\mathbf{x}^{(2)} - \mathbf{x}_k^{(2)}\right) d_k$$

Aligning with non-linear transformations

$$\mathcal{L} = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta \left(\mathbf{x}_n^{(2)} \right) \right\|_2^2, \quad \mathbf{x} \in \mathbb{R}^d$$

When aligning with TPS, the cost function becomes

$$\mathcal{L} = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - \mathbf{x}_n^{(2)} - \sum_{k=1}^{N} \phi \left(\mathbf{x}_n^{(2)} - \mathbf{x}_k^{(2)} \right) d_k \right\|_2^2, \quad \mathbf{x} \in \mathbb{R}^d$$

Note that the displacement vectors are defined through the same landmarks. Similar to linear, the loss is minimized to determine the best transformation parameters **A** and **t**

$$\arg_{\{d_k\}_{k=1,\dots,N}} \min \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - \mathbf{x}_n^{(2)} - \sum_{k=1}^{N} \phi \left(\mathbf{x}_n^{(2)} - \mathbf{x}_k^{(2)} \right) d_k \right\|_2^2$$

Exercise

Can the above minimization's solution be determined in an analytical form? Can you derive it? [Hint: write everything in matrix form.]

Example I - TPS transformation, same subject, same modality

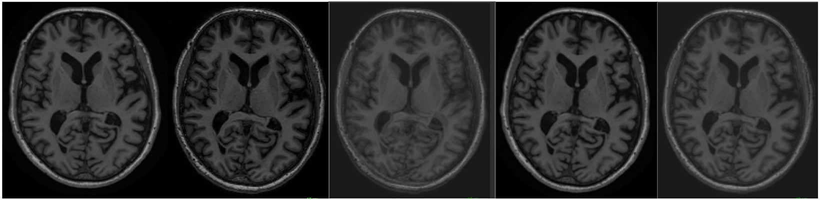


Figure: From left to right: time point #1, time point #2, overlay, time point #1 registered with TPS using 5 landmarks, overlay after registration

- Results from TPS registration - T_θ is a TPS transformation.
- Notice that the overlay is sharper - meaning good alignment

Example II - TPS transformation, different subject, same modality

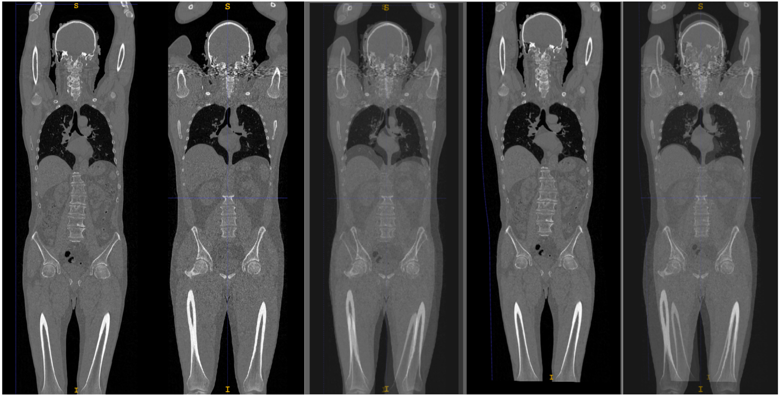
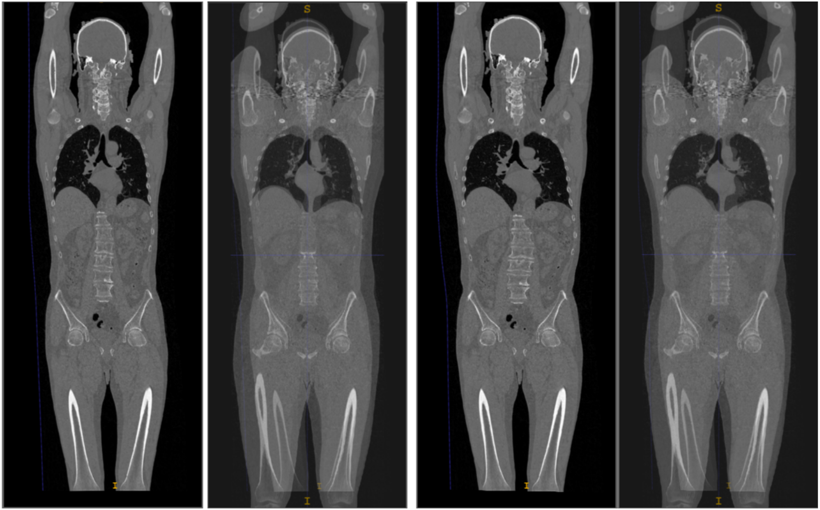


Figure: From left to right: subject #1, subject #2, overlay, subject #1 registered with TPS registration with 12 landmarks, overlay after registration

- TPS has a larger degrees of freedom and model complicated transformations
- 12 different landmarks are used to get a good alignment.
- Notice that landmarks are perfectly aligned but further, the alignment in the rest of the image is also much better than linear transformations.

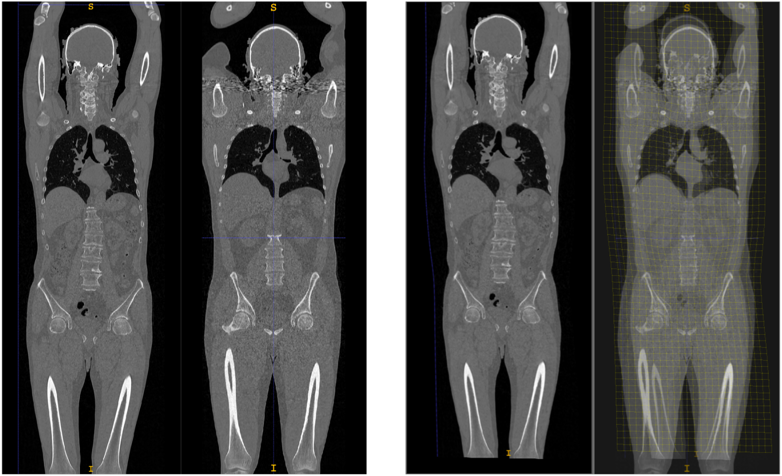
Analysis: Even more landmarks can lead to better alignment in TPS



Seven landmarks

12 landmarks

TPS can model deformable / non-linear transformations



Moving/Source image Fixed/Target image After alignment Overlay and transformation grid

Comparisons between linear and non-linear

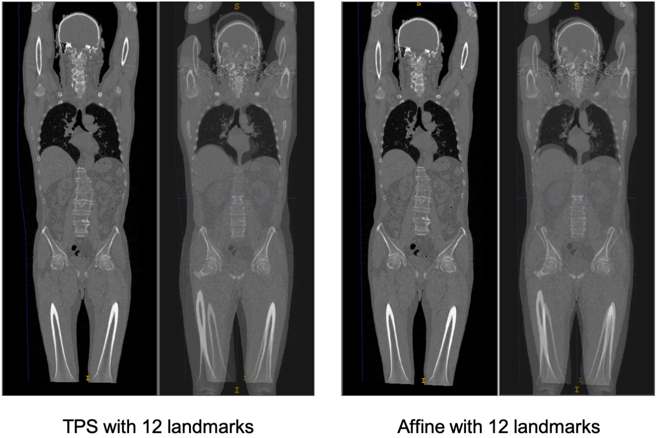


Figure: Usig 12 landmarks for TPS and Affine transformations lead to different results. TPS leads to better alignment.

Further notes on landmark-based registration

- Other transformation models can also be used in the same optimization:

$$\mathcal{L} = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta \left(\mathbf{x}_n^{(2)} \right) \right\|_2^2, \ \mathbf{x} \in \mathbb{R}^d$$

Further notes on landmark-based registration

- Other transformation models can also be used in the same optimization:

$$\mathcal{L} = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta \left(\mathbf{x}_n^{(2)} \right) \right\|_2^2, \ \mathbf{x} \in \mathbb{R}^d$$

- We only considered placing landmarks manually, however, automatic determination of landmarks is also possible with machine learning algorithms.

Further notes on landmark-based registration

- Other transformation models can also be used in the same optimization:

$$\mathcal{L} = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta \left(\mathbf{x}_n^{(2)} \right) \right\|_2^2, \ \mathbf{x} \in \mathbb{R}^d$$

- We only considered placing landmarks manually, however, automatic determination of landmarks is also possible with machine learning algorithms.
- More robust loss functions can be used instead of the ℓ_2 loss. ℓ_1 loss is one such example.
- Robust losses are especially important if we are suspicious about the quality of landmarks.
- For example when you use automatic methods to detect such landmarks.

Further notes on landmark-based registration

- Other transformation models can also be used in the same optimization:

$$\mathcal{L} = \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta\left(\mathbf{x}_n^{(2)}\right) \right\|_2^2, \ \mathbf{x} \in \mathbb{R}^d$$

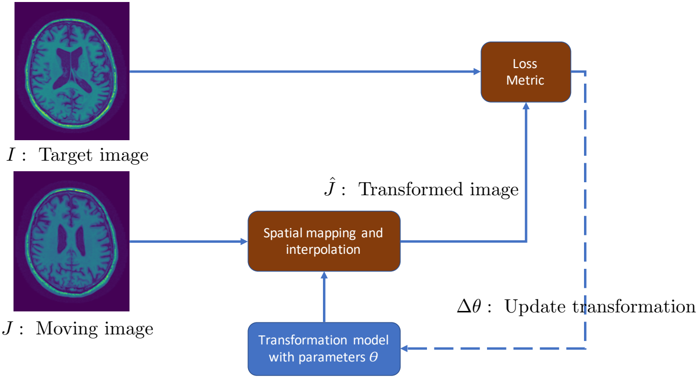
- We only considered placing landmarks manually, however, automatic determination of landmarks is also possible with machine learning algorithms.
- More robust loss functions can be used instead of the ℓ_2 loss. ℓ_1 loss is one such example.
- Robust losses are especially important if we are suspicious about the quality of landmarks.
- For example when you use automatic methods to detect such landmarks.
- Identifying all interest points and then using robust fitting methods, such as RANSAC, is also possible to use. This is particularly used for natural images but less so for medical images.

Outline

Section 4

Intensity-based registration

Overview on a registration algorithm

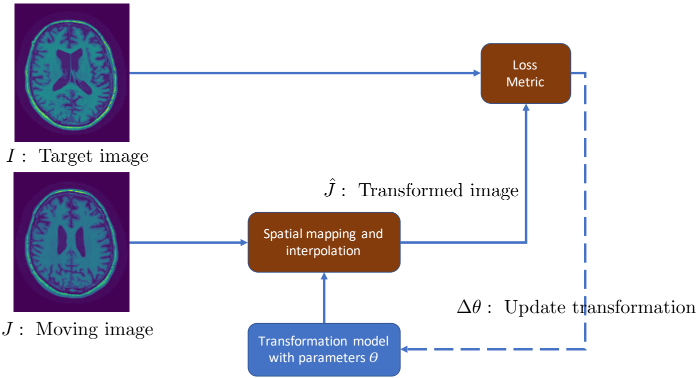


$$\theta^* = \arg_\theta \min \mathcal{L}(\theta)$$

Main questions of today

What loss metrics $\mathcal{L}(\theta)$ do we use and how do we optimize, i.e. solve the $\arg_\theta \min$ problem?

Overview on a registration algorithm



$$\theta^* = \arg_\theta \min \mathcal{L}(\theta)$$

Intensity-based loss function

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J),$$

where $d(\cdot, \cdot)$ is a distance metric between intensities.

Intensity-based registration

CT images of different individuals showing the same FOV.

$$\mathcal{L} = d(I, T_\theta \circ J),$$

- distance $d(\cdot, \cdot)$ is defined over the intentisites of the fixed image I and the moving image J after transformation, i.e. $T_\theta \circ J$.
- there are no landmarks, hence no input information on corresponding locations in the images.
- harder problem as there is no unique solution to the problem.
- definition of $d(\cdot, \cdot)$ is very important and depends on the type of registration.

Intra-modality loss metrics - SSD

One of the most commonly used metrics is the Sum of Squared Distances (SSD)

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J) = \sum_{\mathbf{x} \in \Omega} \| I(\mathbf{x}) - (T_\theta \circ J)(\mathbf{x}) \|_2^2 \,,$$

where Ω is the image domain of the fixed image and $(T_\theta \circ J)(\mathbf{x})$ refers to the intensity of the transformed moving image at $\mathbf{x}$.

Intra-modality loss metrics - SSD

One of the most commonly used metrics is the Sum of Squared Distances (SSD)

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J) = \sum_{\mathbf{x} \in \Omega} \| I(\mathbf{x}) - (T_\theta \circ J)(\mathbf{x}) \|_2^2 ,$$

where Ω is the image domain of the fixed image and $(T_\theta \circ J)(\mathbf{x})$ refers to the intensity of the transformed moving image at $\mathbf{x}$.

Another way of writing the same distance using the explicit transformations is

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J) = \sum_{\mathbf{x} \in \Omega} \left\| I(\mathbf{x}) - J(T_\theta^{-1}(\mathbf{x})) \right\|_2^2 ,$$

SSD's properties

One of the most commonly used metrics is the Sum of Squared Distances (SSD)

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J) = \sum_{\mathbf{x} \in \Omega} \left\| I(\mathbf{x}) - J(T_\theta^{-1}(\mathbf{x})) \right\|_2^2 ,$$

And the registration is defined as

$$\arg_\theta \min \sum_{\mathbf{x} \in \Omega} \left\| I(\mathbf{x}) - J(T_\theta^{-1}(\mathbf{x})) \right\|_2^2 ,$$

SSD's properties

One of the most commonly used metrics is the Sum of Squared Distances (SSD)

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J) = \sum_{\mathbf{x} \in \Omega} \left\| I(\mathbf{x}) - J(T_\theta^{-1}(\mathbf{x})) \right\|_2^2 ,$$

And the registration is defined as

$$\arg_\theta \min \sum_{\mathbf{x} \in \Omega} \left\| I(\mathbf{x}) - J(T_\theta^{-1}(\mathbf{x})) \right\|_2^2 ,$$

- When the images are the same then SSD becomes 0.

SSD's properties

One of the most commonly used metrics is the Sum of Squared Distances (SSD)

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J) = \sum_{\mathbf{x}\in\Omega} \left\| I(\mathbf{x}) - J(T_\theta^{-1}(\mathbf{x})) \right\|_2^2 ,$$

And the registration is defined as

$$\arg_\theta \min \sum_{\mathbf{x}\in\Omega} \left\| I(\mathbf{x}) - J(T_\theta^{-1}(\mathbf{x})) \right\|_2^2 ,$$

- When the images are the same then SSD becomes 0.
- There is no unique solution to this problem. I can warp images in any way to perfectly minimize the SSD to its minimum value 0.

SSD's properties

One of the most commonly used metrics is the Sum of Squared Distances (SSD)

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J) = \sum_{\mathbf{x}\in\Omega} \left\| I(\mathbf{x}) - J(T_\theta^{-1}(\mathbf{x})) \right\|_2^2,$$

And the registration is defined as

$$\arg_\theta \min \sum_{\mathbf{x}\in\Omega} \left\| I(\mathbf{x}) - J(T_\theta^{-1}(\mathbf{x})) \right\|_2^2,$$

- When the images are the same then SSD becomes 0.
- There is no unique solution to this problem. I can warp images in any way to perfectly minimize the SSD to its minimum value 0.
- It assumes the intensities are comparable across the images. Hence, it would not work for registering images with different modalities.

Image gradients are important for SSD

Let us look at the optimization problem

$$\arg_\theta \min \sum_{\mathbf{x}\in\Omega} \left\| I(\mathbf{x}) - J(T_\theta^{-1}(\mathbf{x})) \right\|_2^2,$$

Image gradients are important for SSD

Let us look at the optimization problem

$$\arg_\theta \min \sum_{\mathbf{x}\in\Omega} \left\| I(\mathbf{x}) - J(T_\theta^{-1}(\mathbf{x})) \right\|_2^2,$$

Assume that there is only one parameter θ and we compute the derivative the cost with respect to this parameter

$$\frac{d\mathcal{L}(\theta)}{d\theta} \propto \sum_{\mathbf{x}\in\Omega} \left(I(\mathbf{x}) - J\left(T_\theta^{-1}(\mathbf{x})\right)\right) \nabla J\left(T_\theta^{-1}(\mathbf{x})\right)^T \frac{dT_\theta^{-1}(\mathbf{x})}{d\theta}$$

Image gradients are important for SSD

Let us look at the optimization problem

$$\arg_\theta \min \sum_{\mathbf{x}\in\Omega} \left\| I(\mathbf{x}) - J(T_\theta^{-1}(\mathbf{x})) \right\|_2^2,$$

Assume that there is only one parameter θ and we compute the derivative the cost with respect to this parameter

$$\frac{d\mathcal{L}(\theta)}{d\theta} \propto \sum_{\mathbf{x}\in\Omega} \left(I(\mathbf{x}) - J\left(T_\theta^{-1}(\mathbf{x})\right)\right) \nabla J \left(T_\theta^{-1}(\mathbf{x})\right)^T \frac{dT_\theta^{-1}(\mathbf{x})}{d\theta}$$

Let us focus on the dot product at the far right

$$\nabla J \left(T_\theta^{-1}(\mathbf{x})\right)^T \frac{dT_\theta^{-1}(\mathbf{x})}{d\theta}$$

This term is non-zero when $\frac{dT_\theta^{-1}(\mathbf{x})}{d\theta}$ has a component parallel to the gradient of the transposed image ∇J.

Image gradients are important for SSD

Let us look at the optimization problem

$$\arg_{\theta} \min \sum_{\mathbf{x} \in \Omega} \left\| I(\mathbf{x}) - J(T_{\theta}^{-1}(\mathbf{x})) \right\|_2^2 ,$$

Assume that there is only one parameter θ and we compute the derivative the cost with respect to this parameter

$$\frac{d\mathcal{L}(\theta)}{d\theta} \propto \sum_{\mathbf{x} \in \Omega} \left(I(\mathbf{x}) - J\left(T_{\theta}^{-1}(\mathbf{x})\right)\right) \nabla J \left(T_{\theta}^{-1}(\mathbf{x})\right)^T \frac{dT_{\theta}^{-1}(\mathbf{x})}{d\theta}$$

Let us focus on the dot product at the far right

$$\nabla J \left(T_{\theta}^{-1}(\mathbf{x})\right)^T \frac{dT_{\theta}^{-1}(\mathbf{x})}{d\theta}$$

This term is non-zero when $\frac{dT_{\theta}^{-1}(\mathbf{x})}{d\theta}$ has a component parallel to the gradient of the transposed image ∇J.
Therefore, only changes in transformation that are parallel to the image gradient can change, hence reduce, the loss.

Variations from SSD

Original SSD:

$$d(I, T_\theta \circ J) = \sum_{\mathbf{x} \in \Omega} \left\| I(\mathbf{x}) - J(T_\theta^{-1}(\mathbf{x})) \right\|_2^2$$

One can consider multiple variations on the SSD to improve its performance.

Variations from SSD

Original SSD:

$$d(I, T_\theta \circ J) = \sum_{\mathbf{x}\in\Omega} \left\| I(\mathbf{x}) - J(T_\theta^{-1}(\mathbf{x})) \right\|_2^2$$

One can consider multiple variations on the SSD to improve its performance.

- Using a robust loss instead of ell_2, such as ℓ_1.
- Optimization may become more difficult.

Variations from SSD

Original SSD:

$$d(I, T_\theta \circ J) = \sum_{\mathbf{x}\in\Omega} \left\| I(\mathbf{x}) - J(T_\theta^{-1}(\mathbf{x})) \right\|_2^2$$

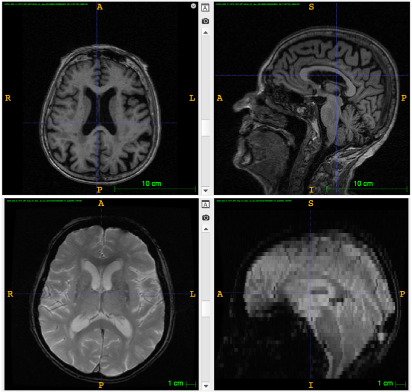
One can consider multiple variations on the SSD to improve its performance.

- Using a robust loss instead of ell_2, such as ℓ_1.
- Optimization may become more difficult.
- Instead of using intensities one can consider features extracted from larger patches

$$d(I, T_\theta \circ J) = \sum_{\mathbf{x}\in\Omega} \left\| f\left(I(\mathcal{N}(\mathbf{x}))\right) - f\left(J(\mathcal{N}(T_\theta^{-1}(\mathbf{x})))\right) \right\|_2^2,$$

where $f()$ is a feature and $\mathcal{N}(\cdot)$ denotes a patch in the image with the argument indicating the center, for instance.

Inter-modality loss metrics



Two different MRI contrasts of the same individual. Top to bottom: Structural and T2star.

- SSD assumes intensities are comparable.
- In inter-modality registration, when we are aligning images from different modalities, we cannot assume they are directly comparable.
- Inter-modality loss metrics are of statistical form.
- Two of the most commonly used ones are
  - Mutual information
  - Normalized cross correlation
- They both assume there is a statistical relationship between intensities of the two images.

Normalized Cross Correlation (NCC)

Normalized cross correlation is defined as

$$\begin{aligned}\mathcal{L}(\theta) &= d(I, T_\theta \circ J) = \mathrm{NCC}(I, J) \\ \mathrm{NCC}(I, J) &= \frac{1}{|\Omega|}\frac{1}{\sigma_I \sigma_J}\sum_{x\in\Omega}(I(\mathbf{x}) - \mu_I)\left(J(T_\theta^{-1}(\mathbf{x})) - \mu_J\right),\end{aligned}$$

where $|\Omega|$ denotes the number of pixels in the fixed image I, σ_I and σ_J are the standard deviations of I and $T_\theta \circ J$, and μ_I and μ_J are the means of the same.

Normalized Cross Correlation (NCC)

Normalized cross correlation is defined as

$$\begin{aligned} \mathcal{L}(\theta) &= d(I, T_\theta \circ J) = \mathrm{NCC}(I, J) \\ \mathrm{NCC}(I, J) &= \frac{1}{|\Omega|} \frac{1}{\sigma_I \sigma_J} \sum_{x \in \Omega} (I(\mathbf{x}) - \mu_I) \left(J(T_\theta^{-1}(\mathbf{x})) - \mu_J \right), \end{aligned}$$

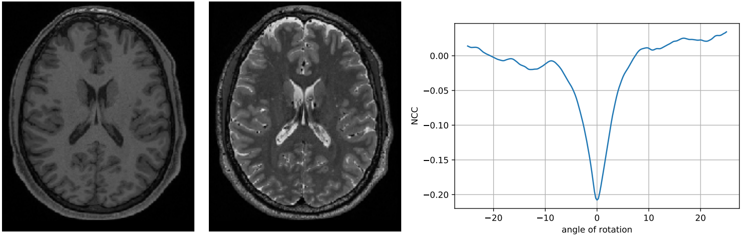
where $|\Omega|$ denotes the number of pixels in the fixed image I, σ_I and σ_J are the standard deviations of I and $T_\theta \circ J$, and μ_I and μ_J are the means of the same.

- NCC is the 2D version of the Pearson correlation coefficient.
- It assumes pixel intensities as independent random variables.
- It achieves maximum score of 1 if $I(\mathbf{x})$ and $J(T_\theta^{-1}(\mathbf{x}))$ are linearly related and the same relationship holds for all $x \in \Omega$.
- It can be thought to approximate

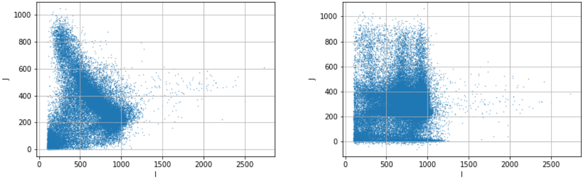
  $$\mathbb{E}\left[(I - \mu_I) * (J - \mu_J)\right] / \sigma_I \sigma_J$$

  if I and J were realy random variables.
- NCC can be both positive and negative. Both cases are fine. It may be better to maximize NCC^2 for registration.

NCC in action

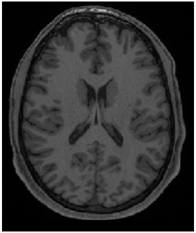


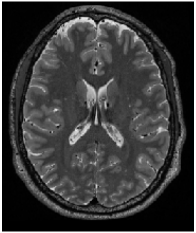
T1w image, T2w image, NCC computed at different angle of rotations. Images are aligned originally. NCC is computed only in the foreground.

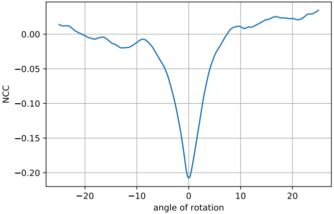


Intensity scatter plots at 0 and 10 degrees of rotation

NCC in action







T1w image, T2w image, NCC computed at different angle of rotations. Images are aligned originally. NCC is computed only in the foreground.

Notice that the NCC does not go down to -1. This is because in reality the relationship is never perfectly linear.

Mutual Information

Mutual information is defined as

$$\begin{aligned}\mathcal{L}(\theta) &= d(I, T_\theta \circ J) = \mathrm{MI}(I, J) \\ \mathrm{MI}(I, J) &= D_{KL}\left(p(I, J) | p(I)p(J)\right)\end{aligned}$$

where $D_{KL}(P|Q)$ is the Kullback-Leibler divergence that measures a distance between two distributions P and Q.
MI measures Kullback-Leibler divergence between the joint distribution of intensities and product of marginal distributions of the same.

Mutual Information

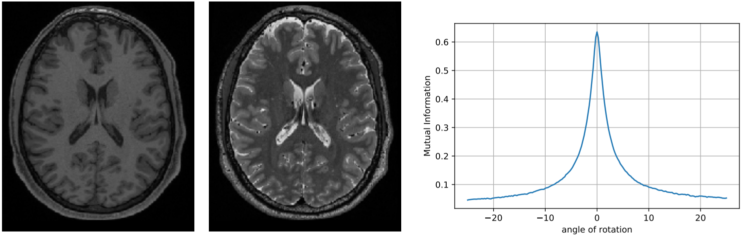
Mutual information is defined as

$$\begin{aligned} \mathcal{L}(\theta) &= d(I, T_\theta \circ J) = \mathrm{MI}(I, J) \\ \mathrm{MI}(I, J) &= D_{KL}\left(p(I, J) | p(I)p(J)\right) \end{aligned}$$

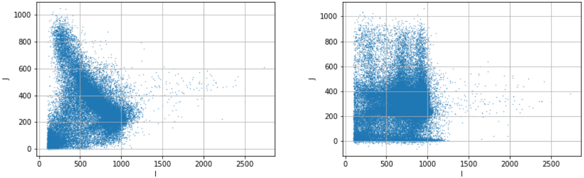
where $D_{KL}(P|Q)$ is the Kullback-Leibler divergence that measures a distance between two distributions P and Q.
MI measures Kullback-Leibler divergence between the joint distribution of intensities and product of marginal distributions of the same.

- It assumes pixel intensities as independent random variables.
- MI takes the value 0 if $p(I, J) = p(I)p(J)$, when the intensities in the two images are independent from each other - no statistical relationship.
- Images that are not aligned would have no statistical relationship, therefore, low MI.
- Perfectly aligned images should have a strong statistical relationship.
- Hence, MI should be maximized during a registration process.

MI in action



T1w image, T2w image, NCC computed at different angle of rotations. Images are aligned originally. NCC is computed only in the foreground.



Intensity scatter plots at 0 and 10 degrees of rotation

Computation of Mutual Information

Mutual information is defined as

$$\begin{aligned} \mathcal{L}(\theta) &= d(I, T_\theta \circ J) = \mathrm{MI}(I, J) \\ \mathrm{MI}(I, J) &= D_{KL}\left(p(I, J) | p(I) p(J)\right) \end{aligned}$$

Computation of Mutual Information

Mutual information is defined as

$$\begin{aligned}\mathcal{L}(\theta) &= d(I, T_\theta \circ J) = \mathrm{MI}(I, J)\\ \mathrm{MI}(I, J) &= D_{KL}\left(p(I, J)|p(I)p(J)\right)\end{aligned}$$

The KL divergence is given in continuous domain as

$$\mathrm{MI}(I, J) = \int_I \int_J p(I, J) \log \frac{p(I, J)}{p(I)p(J)} dIdJ$$

Computation of Mutual Information

Mutual information is defined as

$$\begin{aligned}\mathcal{L}(\theta) &= d(I, T_\theta \circ J) = \mathrm{MI}(I, J) \\ \mathrm{MI}(I, J) &= D_{KL}\left(p(I, J) | p(I)p(J)\right)\end{aligned}$$

The KL divergence is given in continuous domain as

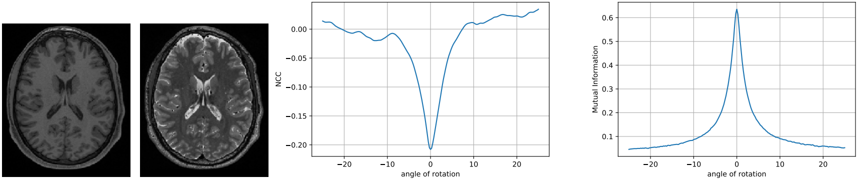
$$\mathrm{MI}(I, J) = \int_I \int_J p(I, J) \log \frac{p(I, J)}{p(I)p(J)} dIdJ$$

We can approximate $p(I)$ and $p(J)$ with histograms of I and $T_\theta \circ J$, respectively. $p(I, J)$ is approximated with 2D histogram of I and $T_\theta \circ J$. With these approximations

$$\mathrm{MI}(I, J) = \sum_i \sum_j h(i, j) \log \frac{h(i, j)}{h(i)h(j)},$$

where h denotes histograms and i, j go through different bins of the histograms.

Comparison of NCC and MI



- MI is more powerful, it can capture complicated relationships.
- For instance, it does not have ripples for large angles in the rotation example.
- MI can be difficult to compute.
- NCC is easier to compute.

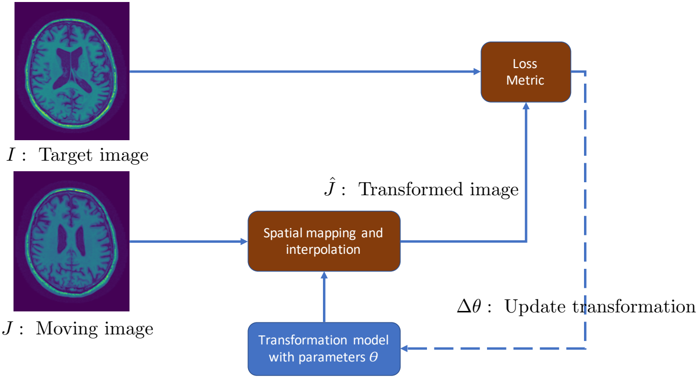
Outline

- Classifying registration problems
- Landmark-based registration
  - Landmarks
  - Linear registration
  - Thin Plate Splines
- Intensity-based registration
  - Intra-modality loss metrics
  - Inter-modality loss metrics
- Optimization - briefly
  - Regularization
  - Gradient descent
  - Sequential optimization

Section 5

Optimization - briefly

Overview on a registration algorithm

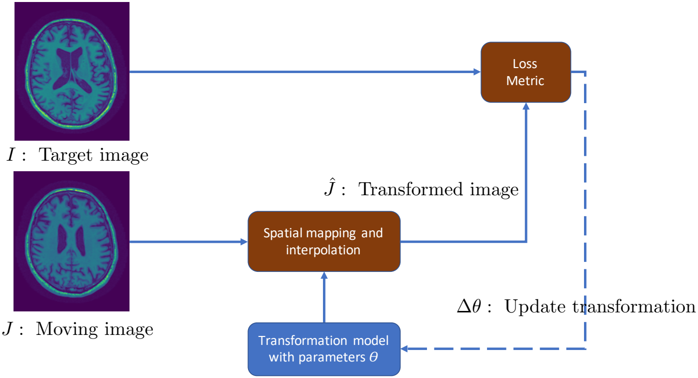


$$\theta^* = \arg_\theta \min \mathcal{L}(\theta)$$

Main questions of today

What loss metrics $\mathcal{L}(\theta)$ do we use and how do we optimize, i.e. solve the $\arg_\theta \min$ problem?

Overview on a registration algorithm



$$\theta^* = \arg_\theta \min \mathcal{L}(\theta)$$

We have seen the spatial mapping, interpolation, transformation parameters and the loss metric. The last remaining block to complete our discussion on registration is the optimization routine where the parameters are updated.

Optimization

We have seen different definitions for $\mathcal{L}(\theta)$ in the lecture

Landmark-based cost function

$$\arg_\theta \min \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta \left(\mathbf{x}_n^{(2)} \right) \right\|_2^2$$

Intensity-based cost function

$$\arg_\theta \min \mathrm{SSD}(I, T_\theta \circ J)$$
$$\arg_\theta \min \mathrm{NCC}(I, T_\theta \circ J)$$
$$\arg_\theta \min \mathrm{MI}(I, T_\theta \circ J)$$

Optimization

We have seen different definitions for $\mathcal{L}(\theta)$ in the lecture

Landmark-based cost function

$$\arg_\theta \min \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta \left(\mathbf{x}_n^{(2)} \right) \right\|_2^2$$

Intensity-based cost function

$$\arg_\theta \min \mathrm{SSD}(I, T_\theta \circ J)$$
$$\arg_\theta \min \mathrm{NCC}(I, T_\theta \circ J)$$
$$\arg_\theta \min \mathrm{MI}(I, T_\theta \circ J)$$

- Optimization for image registration is a difficult problem in general.

Optimization

We have seen different definitions for $\mathcal{L}(\theta)$ in the lecture

Landmark-based cost function

$$\arg_\theta \min \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta\left(\mathbf{x}_n^{(2)}\right) \right\|_2^2$$

Intensity-based cost function

$$\arg_\theta \min \mathrm{SSD}(I, T_\theta \circ J)$$
$$\arg_\theta \min \mathrm{NCC}(I, T_\theta \circ J)$$
$$\arg_\theta \min \mathrm{MI}(I, T_\theta \circ J)$$

- Optimization for image registration is a difficult problem in general.
- Landmark-based cost function is easier of the two. As long as degrees of freedom of the transformation is consistent with the given number of landmarks.

Optimization

We have seen different definitions for $\mathcal{L}(\theta)$ in the lecture

Landmark-based cost function

$$\arg_\theta \min \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta\left(\mathbf{x}_n^{(2)}\right) \right\|_2^2$$

Intensity-based cost function

$$\arg_\theta \min \mathrm{SSD}(I, T_\theta \circ J)$$
$$\arg_\theta \min \mathrm{NCC}(I, T_\theta \circ J)$$
$$\arg_\theta \min \mathrm{MI}(I, T_\theta \circ J)$$

- Optimization for image registration is a difficult problem in general.
- Landmark-based cost function is easier of the two. As long as degrees of freedom of the transformation is consistent with the given number of landmarks.
- Optimization for linear transformation models is easier than non-linear transformation models.

Optimization

We have seen different definitions for $\mathcal{L}(\theta)$ in the lecture

Landmark-based cost function

$$\arg_\theta \min \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta\left(\mathbf{x}_n^{(2)}\right) \right\|_2^2$$

Intensity-based cost function

$$\arg_\theta \min \mathrm{SSD}(I, T_\theta \circ J)$$
$$\arg_\theta \min \mathrm{NCC}(I, T_\theta \circ J)$$
$$\arg_\theta \min \mathrm{MI}(I, T_\theta \circ J)$$

- Optimization for image registration is a difficult problem in general.
- Landmark-based cost function is easier of the two. As long as degrees of freedom of the transformation is consistent with the given number of landmarks.
- Optimization for linear transformation models is easier than non-linear transformation models.
- In general, intensity-based registration problems using non-linear transformation models (named as deformable registration) are ill-posed, i.e. there are many different solutions that can give similar results.

Optimization

We have seen different definitions for $\mathcal{L}(\theta)$ in the lecture

Landmark-based cost function

$$\arg_\theta \min \sum_{n=1}^{N} \left\| \mathbf{x}_n^{(1)} - T_\theta\left(\mathbf{x}_n^{(2)}\right) \right\|_2^2$$

Intensity-based cost function

$$\arg_\theta \min \mathrm{SSD}(I, T_\theta \circ J)$$
$$\arg_\theta \min \mathrm{NCC}(I, T_\theta \circ J)$$
$$\arg_\theta \min \mathrm{MI}(I, T_\theta \circ J)$$

- Optimization for image registration is a difficult problem in general.
- Landmark-based cost function is easier of the two. As long as degrees of freedom of the transformation is consistent with the given number of landmarks.
- Optimization for linear transformation models is easier than non-linear transformation models.
- In general, intensity-based registration problems using non-linear transformation models (named as deformable registration) are ill-posed, i.e. there are many different solutions that can give similar results.
- To overcome the ill-posed nature of the problem, regularization is often used when using non-linear transformation models.

Completing the cost function with continuous regularization models

To make the optimization well-posed, a regularization is often added to the loss functions

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J) + \lambda r(T_\theta), \;\; T_\theta = \mathbf{x} + u(\mathbf{x})$$

where $r(T_\theta)$ prefers some sort of transformations over others and λ is a weight.

Completing the cost function with continuous regularization models

To make the optimization well-posed, a regularization is often added to the loss functions

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J) + \lambda r(T_\theta), \;\; T_\theta = \mathbf{x} + u(\mathbf{x})$$

where $r(T_\theta)$ prefers some sort of transformations over others and λ is a weight. Tikhonov regularization schemes that prefer smoothly varying $u(\mathbf{x})$ are often used. The generic form is

$$r(T_\theta) = \int_\Omega \phi(\nabla u) d\mathbf{x},$$

where ∇ denotes gradient in space and ϕ is a convex function with minimum at zero. Examples are

Completing the cost function with continuous regularization models

To make the optimization well-posed, a regularization is often added to the loss functions

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J) + \lambda r(T_\theta), \;\; T_\theta = \mathbf{x} + u(\mathbf{x})$$

where $r(T_\theta)$ prefers some sort of transformations over others and λ is a weight. Tikhonov regularization schemes that prefer smoothly varying $u(\mathbf{x})$ are often used. The generic form is

$$r(T_\theta) = \int_\Omega \phi(\nabla u) d\mathbf{x},$$

where ∇ denotes gradient in space and ϕ is a convex function with minimum at zero. Examples are

- ℓ_2 regularization that prefers smoothly varying displacement fields

$$r(T_\theta) = \int_\Omega (\nabla u)^2 d\mathbf{x}$$

Completing the cost function with continuous regularization models

To make the optimization well-posed, a regularization is often added to the loss functions

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J) + \lambda r(T_\theta), \;\; T_\theta = \mathbf{x} + u(\mathbf{x})$$

where $r(T_\theta)$ prefers some sort of transformations over others and λ is a weight. Tikhonov regularization schemes that prefer smoothly varying $u(\mathbf{x})$ are often used. The generic form is

$$r(T_\theta) = \int_\Omega \phi(\nabla u) d\mathbf{x},$$

where ∇ denotes gradient in space and ϕ is a convex function with minimum at zero. Examples are

- ℓ_2 regularization that prefers smoothly varying displacement fields

$$r(T_\theta) = \int_\Omega (\nabla u)^2 d\mathbf{x}$$

- ℓ_1 regularization that prefers piece-wise constant $u(x)$

$$r(T_\theta) = \int_\Omega |\nabla u| d\mathbf{x}$$

Regularization models from continuum mechanics

To make the optimization well-posed, a regularization is often added to the loss functions

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J) + \lambda r(T_\theta), \;\; T_\theta = \mathbf{x} + u(\mathbf{x})$$

where $r(T_\theta)$ prefers some sort of transformations over others.
Energy terms from continuum mechanics are also commonly used

Regularization models from continuum mechanics

To make the optimization well-posed, a regularization is often added to the loss functions

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J) + \lambda r(T_\theta), \;\; T_\theta = \mathbf{x} + u(\mathbf{x})$$

where $r(T_\theta)$ prefers some sort of transformations over others.
Energy terms from continuum mechanics are also commonly used

- Linear elastic regularization uses the Cauchy strain tensor $V = V(u)$ and minimizes the strain energy caused by the displacement field

$$r(T_\theta) = \int_\Omega \lambda Tr(V)^2 + \mu Tr(V^2) d\mathbf{x}, \;\; V(u) = (\nabla u + \nabla u^T)/2,$$

  where ν and μ are Lame constants and ∇ is the gradient in space. This model effectively prefers small $u(\mathbf{x})$ values.

Regularization models from continuum mechanics

To make the optimization well-posed, a regularization is often added to the loss functions

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J) + \lambda r(T_\theta), \;\; T_\theta = \mathbf{x} + u(\mathbf{x})$$

where $r(T_\theta)$ prefers some sort of transformations over others.
Energy terms from continuum mechanics are also commonly used

- Linear elastic regularization uses the Cauchy strain tensor $V = V(u)$ and minimizes the strain energy caused by the displacement field

$$r(T_\theta) = \int_\Omega \lambda Tr(V)^2 + \mu Tr(V^2) d\mathbf{x}, \;\; V(u) = (\nabla u + \nabla u^T)/2,$$

  where ν and μ are Lame constants and ∇ is the gradient in space. This model effectively prefers small $u(\mathbf{x})$ values.

- Hyperelastic regularization allows for larger deformations using the simplest material model, the Saint Venant–Kirchhoff model with the Green-St.Venan strain tensor $E = E(u)$

$$r(T_\theta) = \int_\Omega \frac{\lambda}{2} Tr(E)^2 + \mu Tr(E^2) d\mathbf{x}, \; E(u) = \frac{1}{2} \nabla T_\theta^T \nabla T_\theta - \mathbb{I} = \frac{1}{2}[\nabla u^T + \nabla u + \nabla u^T \nabla u],$$

  which prefers locally rigid transformations.

Complete registration model is ready for optimization

The complete registration loss function, together with the data term $d()$ and the regularization $r()$ is ready for optimization.
Linear transformation models - Linear registration problem:

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J), \;\; T_\theta = \mathbf{Ax} + \mathbf{t}$$

Non-linear transformation models - non-linear registration problem:

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J) + \lambda r(T_\theta), \;\; T_\theta = \mathbf{x} + u(\mathbf{x})$$

Complete registration model is ready for optimization

The complete registration loss function, together with the data term $d()$ and the regularization $r()$ is ready for optimization.
Linear transformation models - Linear registration problem:

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J), \;\; T_\theta = \mathbf{Ax} + \mathbf{t}$$

Non-linear transformation models - non-linear registration problem:

$$\mathcal{L}(\theta) = d(I, T_\theta \circ J) + \lambda r(T_\theta), \;\; T_\theta = \mathbf{x} + u(\mathbf{x})$$

Most commonly, the optimization can be done via continuous methods

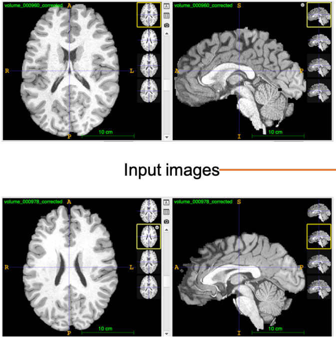
$$\theta_{t+1} = \theta_t + \alpha g(\mathcal{L}(\theta_t)),$$

where α is a step-size parameter, t indices optimization iterations and $g()$ is a gradient-based function, the simplest being the gradient descent

$$\theta_{t+1} = \theta_t - \alpha \frac{d\mathcal{L}(\theta)}{d\theta}|_{\theta_t},$$

Second order optimization schemes and discrete formulations of the optimization is also possible.

Last notes: 1 - Sequential optimization - first linear then deformable



Input images

Linear registration

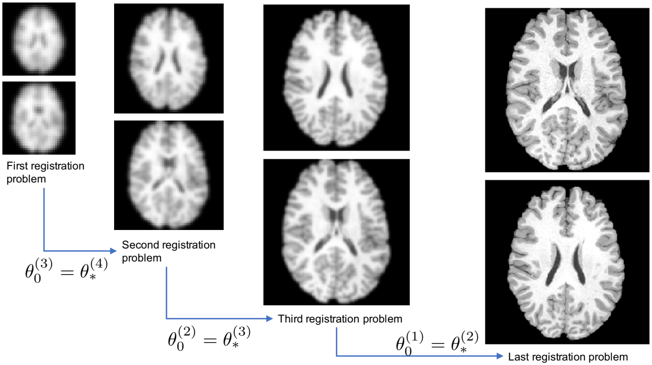
- Get rids of the global misalignment.
- Easier problem to solve.
- Yields a good initialization for the following nonlinear registration problem.

Non-linear / Deformable registration

- Performs the finer alignment.
- It can work well with good initialization.

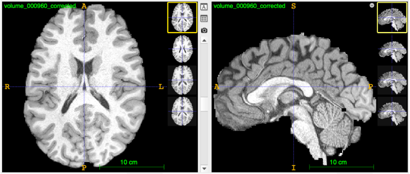
It is common to first align images with linear registration followed by deformable non-linear registration. This helps to get a better alignment.

Last notes: 2 - Multi-scale optimization

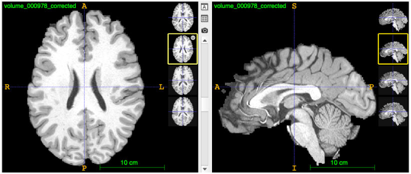


Even with regularization terms, registration can get stuck in local minima. To avoid this mult-scale optimization is used. At coarse resolution, registration is easier because there are not that many details in the images. Registration starts at the coarsest scale where the problem of aligning the images is easier. Results are used as initialization of the next step.

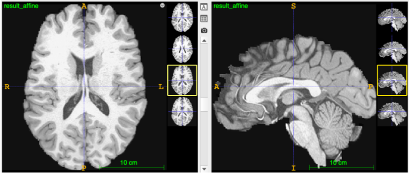
Demonstration



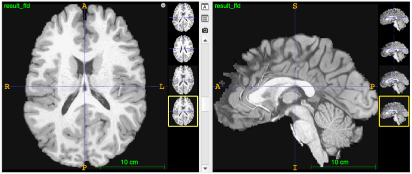
Fixed/Target Image



Moving/Source Image



Affine registration result



FFD with BSplines after affine

Toolboxes

It is a laborious task to code an entire registration tool. Some tools are already available:

- ITK is a C++ library with available registration tools.
- Elastix `http://elastix.isi.uu.nl` is a great and easy to use tool.
- Advanced Normalization Tools (ANTS) `http://stnava.github.io/ANTs/` is another great tool that is easy to use.