Mutual-Information Medical Image Registration: Theory and Examples

Xinzhe Luo

SDS@Fudan University

December 19, 2021

Contents

Information

Information

Given a discrete random variable X with probability distribution $p(x)$, its information is defined as

$$h(x) = -\log p(x),$$

which satisfies the two conditions:

1. The information of an event is inversely related to the probability it takes place;
2. If two events are unrelated, then the information gain from observing both of them should be the sum of the information gained from each of them separately.

Entropy

Shannon's entropy

Given events $e_1, \ldots, e_m$ occurring with probabilities $p_1, \ldots, p_m$, the Shannon's entropy is defined as

$$H(X) = \sum_i p_i \log \frac{1}{p_i} = -\sum_i p_i \log p_i .$$

Remark

The Shannon's entropy has three interpretations:

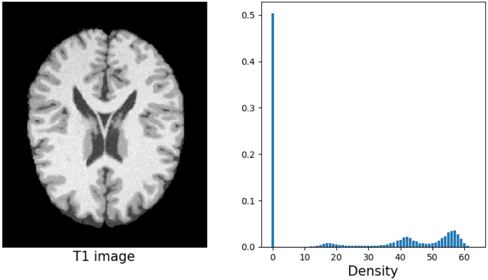
1. *The amount of average information,*
2. *The uncertainty of the random variable,*
3. *The dispersion of the probability distribution.*

Intensity Distribution

Histogramming

The intensity distribution of an image $I : \Omega \to \mathbb{R}$ can be given by a binning function $B : \mathbb{R} \to \{1, \ldots, n\}$ and the discrete r.v. $X = B \circ I$. Then $p_i = P(X = i)$ defines the intensity distribution of image I.

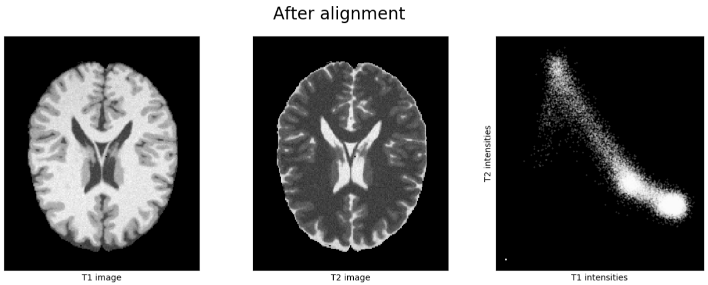
Histogram



Joint Intensity Distribution

Joint histogram

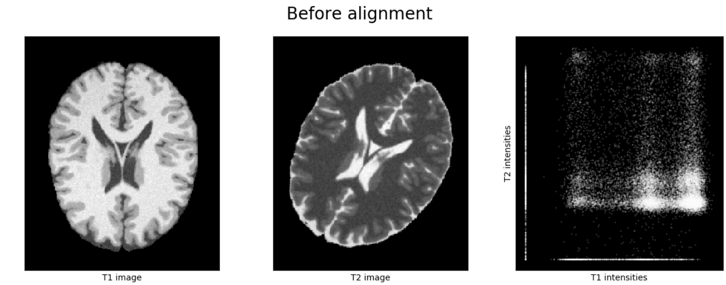
When the images are correctly registered, the joint histogram shows certain clusters for the gray values of anatomical structures.



Joint Intensity Distribution

Joint histogram

As the images become misaligned, the joint intensity histogram displays a dispersion of the clustering.



Joint Entropy

Joint entropy

The dispersion can be measured by the joint entropy of the two images:

$$H(X,Y) = -\sum_i \sum_j P(X=i, Y=j) \log P(X=i, Y=j).$$

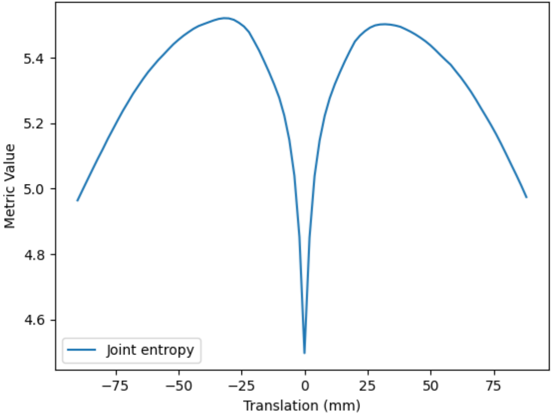
By finding the transformation that minimizes the joint entropy, images should be registered.

Related work

- Collignon, André, et al. "3D multi-modality medical image registration using feature space clustering." CVRMed 1995.
- Studholme, Colin, Derek LG Hill, and David J. Hawkes. "Multiresolution voxel similarity measures for MR-PET registration." IPMI 1995.

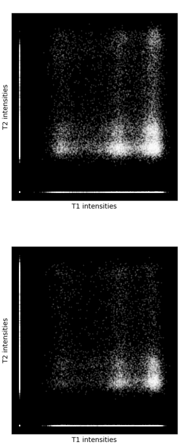
Joint Entropy

Metric simulation



Joint Entropy

Joint histogram



Joint Entropy

Limitation of joint entropy

The capture range of the joint entropy can be very small. When large misalignment occurs, the joint histogram can still be very sharp, leading to decreased values of the joint entropy.

Equivalent Definitions of Mutual Information

Definition 1

For two images X and Y, the mutual information I can be defined as

$$I(X, Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X),$$

which can be interpreted as the reduction of the amount of uncertainty about one image when the other one is known.

Remark

$I(X, Y) \leq H(X)$, *the maximum attains when X is a function of Y almost surely.*

Equivalent Definitions of Mutual Information

Definition 2

MI is related to the joint entropy in the sense

$$I(X, Y) = H(X) + H(Y) - H(X, Y).$$

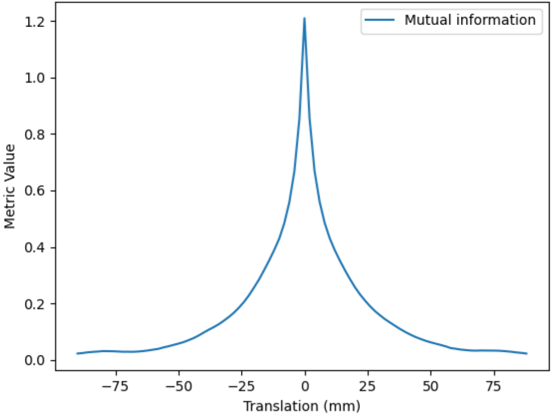
Remark (The advantange of MI over JE)

"The marginal entropies will have low values when the overlapping part of the images contains only background and high values when it contains anatomical structures."

Pluim, Josien PW, JB Antoine Maintz, and Max A. Viergever. "Mutual-information-based registration of medical images: a survey." IEEE transactions on medical imaging 22.8 (2003): 986-1004.

Equivalent Definitions of Mutual Information

Metric simulation



Equivalent Definitions of Mutual Information

Definition 3

MI is related to the Kullback-Leibler divergence in the sense

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

which is a measure of dependence between two images.

Remark

"The assumption is that there is maximal dependence between the gray values of the images when they are correctly aligned."

Pluim, Josien PW, JB Antoine Maintz, and Max A. Viergever. "Mutual-information-based registration of medical images: a survey." IEEE transactions on medical imaging 22.8 (2003): 986-1004.

Properties of Mutual Information

Basic properties

1. Symmetry: $I(X, Y) = I(Y, X)$.
2. Self information: $I(X, X) = H(X)$.
3. Boundedness: $I(X, Y) \leq \min\{H(X), H(Y)\}$.
4. Non-negativity: $I(X, Y) \geq 0$.
5. Independence: $I(X, Y) = 0$ if and only if X and Y are independent.

Remark

More properties can be found in the book: Cover, Thomas M. Elements of information theory. John Wiley & Sons, 1999.

Contents

Framework of Image Registration

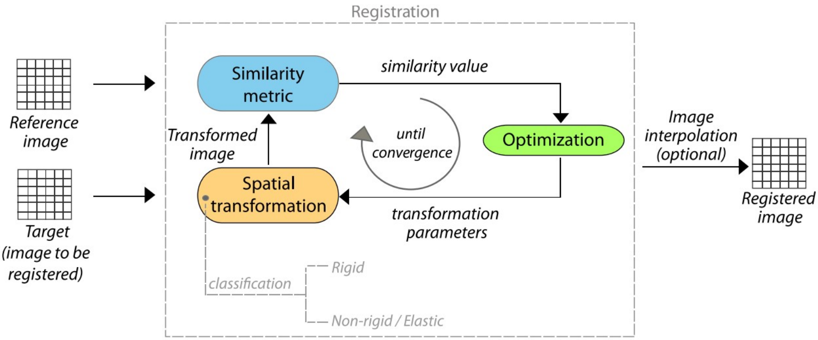


Fig. 4. A typical registration scheme.

[0] Gupta, Vikas, et al. "Cardiac MR perfusion image processing techniques: a survey." Medical image analysis 16.4 (2012): 767-785.

Framework of Image Registration

Mathematical formulation

Given the reference (fixed) image $r : \Omega_r \to \mathbb{R}$, the floating (moving) image $f : \Omega_f \to \mathbb{R}$ and a similarity measure $S(\cdot, \cdot)$, the transformation $T_\alpha : \Omega_r \to \Omega_f$ (parameterized by α) that registers the two images is given by

$$\widehat{\alpha} = \arg\min_{\alpha} S(r, f \circ T_\alpha),$$

where $\circ$ denotes interpolation.

Estimating Mutual Information

Estimating MI

MI entails estimation of the joint intensity distribution, which is often done by calculating the joint histogram

$$h_\alpha(r, f \circ T_\alpha).$$

Problem with histogramming

Histogramming is a discrete operation, making the similarity measure a discrete function of the transformation parameters. The optimization becomes difficult.

Estimating Mutual Information

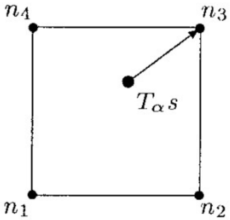
Two approaches

1. Partial volume interpolation: Maes, Frederik, et al. "Multimodality image registration by maximization of mutual information." IEEE transactions on Medical Imaging 16.2 (1997): 187-198.
2. Parzen window estimation or kernel density estimator: Wells III, William M., et al. "Multi-modal volume registration by maximization of mutual information." Medical image analysis 1.1 (1996): 35-51.

Basic Interpolations

Nearest-neighbour interpolation

NN:



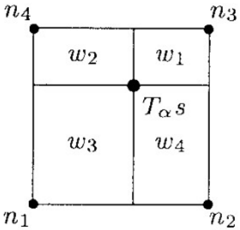
$$\arg\min_{n_i} d(T_\alpha s, n_i) = n_3$$
$$r(T_\alpha s) = r(n_3)$$
$$h_\alpha(f(s), r(T_\alpha s)) \mathrel{+}= 1$$

Maes, Frederik, et al. "Multimodality image registration by maximization of mutual information." IEEE transactions on Medical Imaging 16.2 (1997): 187-198.

Basic Interpolations

Trilinear interpolation

TRI:



$$\sum_i w_i(T_\alpha s) = 1$$
$$r(T_\alpha s) = \sum_i w_i \cdot r(n_i)$$
$$h_\alpha(f(s), r(T_\alpha s)) \mathrel{+}= 1$$

Maes, Frederik, et al. "Multimodality image registration by maximization of mutual information." IEEE transactions on Medical Imaging 16.2 (1997): 187-198.

Basic Interpolations

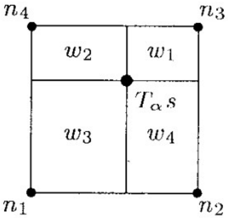
Problem with basic interpolations

Both nearest-neighbour and trilinear interpolation are non-differentiable with respect to the transformation parameters, making gradient-based optimization intractable.

Partial Volume Interpolation

Partial volume interpolation

PV:



$$\sum_i w_i(T_\alpha s) = 1$$
$$\forall i : h_\alpha(f(s), r(n_i)) \mathrel{+}= w_i$$

Maes, Frederik, et al. "Multimodality image registration by maximization of mutual information." IEEE transactions on Medical Imaging 16.2 (1997): 187-198.

Parzen Window Estimation

Estimating probability density function

The Parzen window estimation of a probability density function is given by

$$\mathcal{H}_\alpha(\mu,\nu) = \sum_{s\in\Omega_r} K_r\left(\frac{r(s)-\mu}{h_r}\right) K_f\left(\frac{f\circ T_\alpha(s)-\nu}{h_f}\right),$$

$$p_\alpha(\mu,\nu) = \frac{\mathcal{H}_\alpha(\mu,\nu)}{\sum_{\mu',\nu'}\mathcal{H}_\alpha(\mu',\nu')}.$$

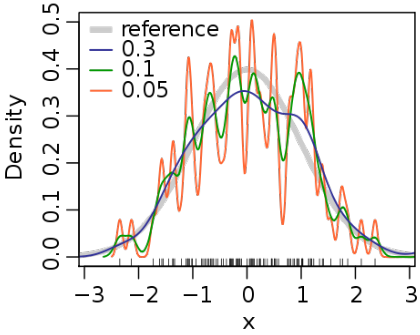
where K is the kernel satisfying $\int K(u)\,\mathrm{d}u = 1$ and h is the bandwidth of the kernel.

Remark

The Parzen window estimation approximates the true joint intensity distribution by a continuous density function differentiable to transformation parameters.

Parzen Window Estimation

Example



https://en.wikipedia.org/wiki/Kernel_density_estimation#/media/File:Kernel_density.svg

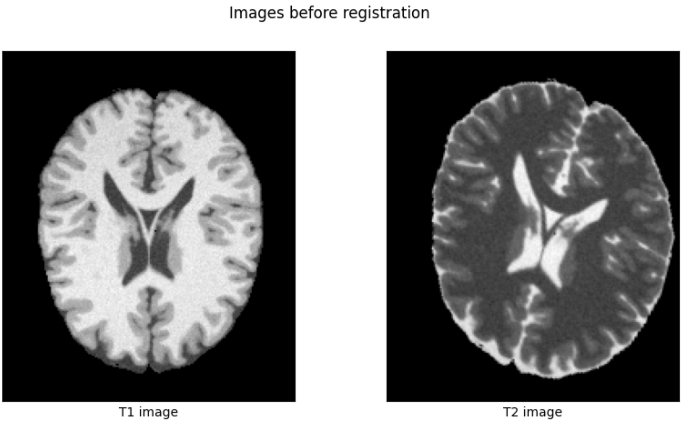
Parzen Window Estimation

Related work

Thévenaz, Philippe, and Michael Unser. "Optimization of mutual information for multiresolution image registration." IEEE transactions on image processing 9.12 (2000): 2083-2099.

Demonstration

Misaligned images



Demonstration

Transformation model

Multi-level registration:

1. Translation;
2. Rigid;
3. Affine.

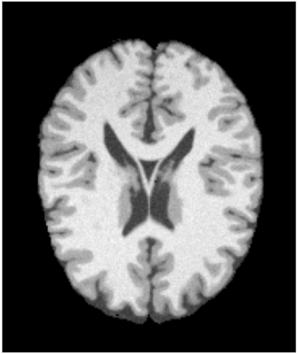
Evaluation

A is the initial affine matrix that causes the two images to be misaligned; $\widehat{A}$ is the affine matrix that registers the misaligned images. The Frobenius norm of $A\widehat{A} - I$ is used to evaluate the registration accuracy.

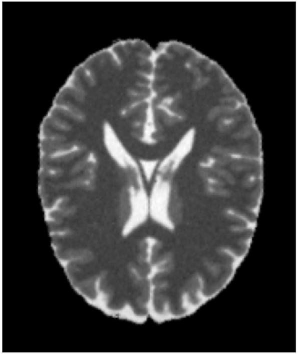
Demonstration

Registered images

Images after registration steps 999



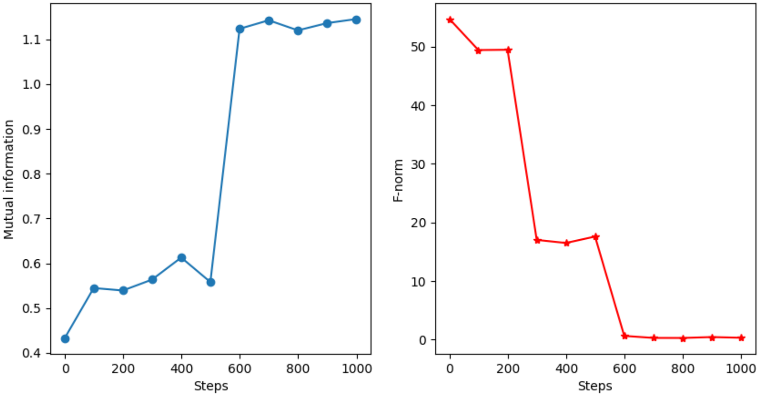
T1 image



T2 image

Demonstration

Evaluation



Contents

Preliminaries

Notations

1. The reference image $r : \Omega_r \to \mathbb{R}$;
2. The floating image $f : \Omega_f \to \mathbb{R}$;
3. The transformation model $T_\alpha : \Omega_r \to \Omega_f$ parameterized by α.

Maximum likelihood estimation

Given an i.i.d. sample $\boldsymbol{X}$, a maximum likelihood estimator of the parameter θ based on $\boldsymbol{X}$ is the parameter value at which the likelihood function $L(\theta \mid \boldsymbol{X})$ attains its maximum as a function of θ:

$$\widehat{\theta}(\boldsymbol{X}) = \arg\max L(\theta \mid \boldsymbol{X}).$$

Maximum Likelihood Registration

Likelihood function

Assuming the intensities of the reference and floating images are corresponded by $\{(r(s), f \circ T_\alpha(s)) : s \in \Omega_r, T_\alpha(s) \in \Omega_f\}$ and independent spatial locations, the likelihood function is written as

$$L(\theta \mid r, f) = \prod_{s \in \Omega_r^\alpha} p(r(s), f \circ T_\alpha(s); \theta)$$

Log-likelihood

The log-likelihood function can then be written as

$$\ell(\theta \mid r, f) = \sum_{s \in \Omega_r^\alpha} \ln p(r(s), f \circ T_\alpha(s); \theta).$$

Maxmimum Likelihood Registration

Categorical models

Assume the images have discrete intensities, and the model is jointly categorical, i.e.

$$p(r(s) = i, f \circ T_\alpha(s) = j; \alpha) = \beta_{ij} \geq 0, \quad \forall\, s \in \Omega_r^\alpha,$$

such that $\sum_{i,j} \beta_{ij} = 1$.

Log-likelihood

Then the likelihood function becomes

$$\ell(\theta \mid r, f) = \sum_{i,j} N_{ij}(\alpha) \ln \beta_{ij},$$

where $N_{ij}(\alpha) \triangleq \#\{s \in \Omega_r^\alpha : r(s) = i, f \circ T_\alpha(s) = j\}$ and $\theta = \{\alpha, \beta\}$.

Maximum Likelihood Registration

Maximum profile likelihood

The transformation parameters α are of primary interest while the model parameters $\{\beta_{ij}\}$ can be viewed as nuisance parameters. Maximizing out the nuisance parameters is called maximum profile likelihood:

$$\begin{aligned}\widehat{\alpha} &= \arg\max_{\alpha} \max_{\beta} \ell(\theta \mid r, f) \\ &= \arg\max_{\alpha} \max_{\beta} \sum_{i,j} N_{ij}(\alpha) \ln \beta_{ij}, \quad \text{s.t.} \sum_{i,j} \beta_{ij} = 1.\end{aligned}$$

Maximum Likelihood Registration

Solving by Lagrange multipliers

$$\widehat{\beta_{ij}}(\alpha) = \frac{N_{ij}(\alpha)}{N(\alpha)} = \widehat{p}(i, j; \alpha),$$

where $N(\alpha) \triangleq |\Omega_r^\alpha|$ and $\widehat{p}(i, j; \alpha)$ is the normalized joint histogram of the two images in the overlap region.

Maximum Likelihood Registration

Relation to minimum joint entropy

Maximizing out β gives

$$\begin{aligned}\widehat{\alpha} &= \arg\max_{\alpha} N(\alpha) \sum_{i,j} \widehat{p}(i,j;\alpha) \ln \widehat{p}(i,j;\alpha) \\ &= \arg\min_{\alpha} N(\alpha) H(r, f \circ T_{\alpha}).\end{aligned}$$

Contents

Overlap Region

The effect of overlap region on similarity measures

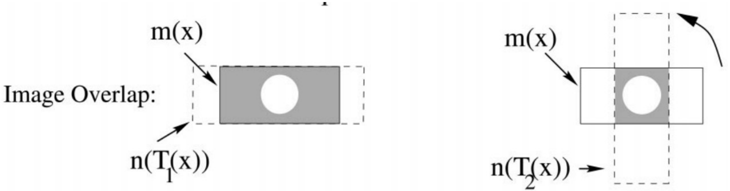


Figure: Left: joint entropy; Right: mutual information.

Studholme, Colin, Derek LG Hill, and David J. Hawkes. "An overlap invariant entropy measure of 3D medical image alignment." Pattern recognition 32.1 (1999): 71-86.

Simulation

Response of similarity measures to overlap region

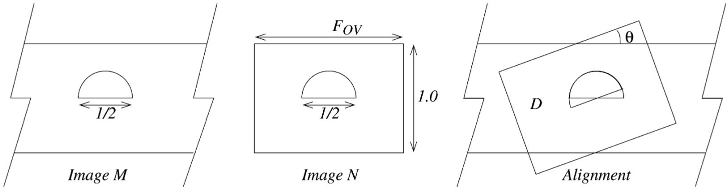


Fig. 6. A simple model of rotational alignment θ between two images of a half circle with varying overlap and horizontal field of view determined by F_{OV}.

Studholme, Colin, Derek LG Hill, and David J. Hawkes. "An overlap invariant entropy measure of 3D medical image alignment." Pattern recognition 32.1 (1999): 71-86.

Simulation

Response of negated joint entropy to overlap region

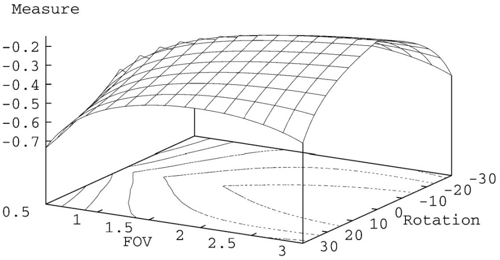


Fig. 7. The response of negated Joint Entropy to rotational misalignment (degrees) at different values of field of view parameter F_{OV} of the model in Fig. 6.

Studholme, Colin, Derek LG Hill, and David J. Hawkes. "An overlap invariant entropy measure of 3D medical image alignment." Pattern recognition 32.1 (1999): 71-86.

Simulation

Response of mutual information to overlap region

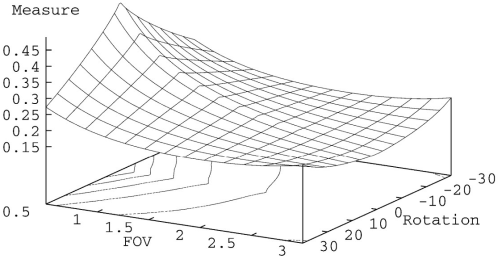


Fig. 8. The response of Mutual Information $I(M:N)$ to rotational misalignment (degrees) at different values of field of view parameter F_{OV} of the model in Fig. 6

Studholme, Colin, Derek LG Hill, and David J. Hawkes. "An overlap invariant entropy measure of 3D medical image alignment." Pattern recognition 32.1 (1999): 71-86.

Normalized Mutual Information

Nomalized mutual information

The normalized mutual information is defined as

$$\mathrm{NMI}(X, Y) = \frac{H(X) + H(Y)}{H(X, Y)}.$$

Simulation

Response of normalized mutual information to overlap region

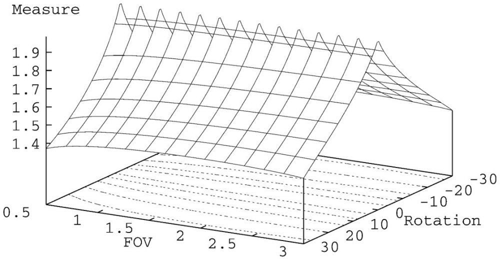
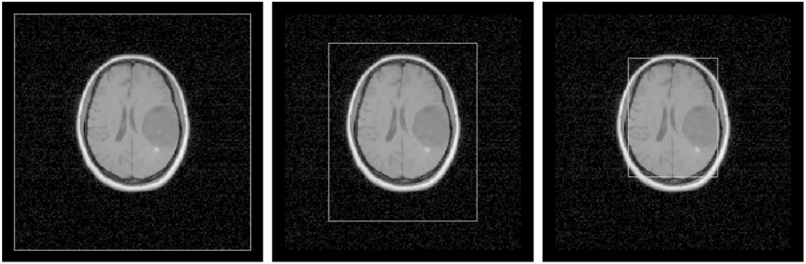


Fig. 9. The response of Normalised Mutual Information $Y(M;N)$ to rotational misalignment (degrees) at different values of field of view parameter F_{OV} of the model in Fig. 6.

Studholme, Colin, Derek LG Hill, and David J. Hawkes. "An overlap invariant entropy measure of 3D medical image alignment." Pattern recognition 32.1 (1999): 71-86.

Experiment

Data



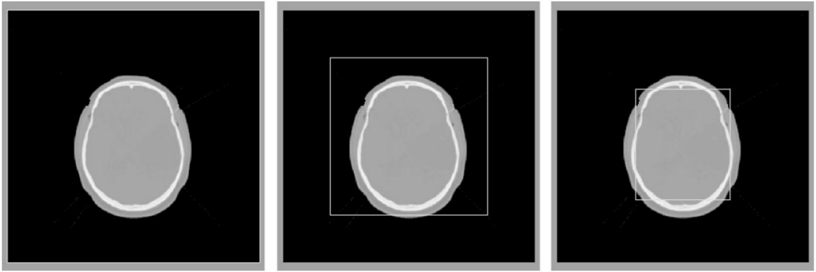
TF1: 320 × 320mm TF2: 200 × 240mm TF3: 120 × 160mm

Fig. 10. The three levels of MR Transaxial field of view used for experiments.

Studholme, Colin, Derek LG Hill, and David J. Hawkes. "An overlap invariant entropy measure of 3D medical image alignment." Pattern recognition 32.1 (1999): 71-86.

Experiment

Data



TF1: 334 × 334mm TF2: × 209mm TF3: 125 × 146mm

Fig. 11. The three levels of CT Transaxial field of view used for experiments.

Studholme, Colin, Derek LG Hill, and David J. Hawkes. "An overlap invariant entropy measure of 3D medical image alignment." Pattern recognition 32.1 (1999): 71-86.

Experiment

Results

Table 1
Number of successfully recovered alignments (as defined in the text) from 50 random starts at three levels of misalignment, as transaxial field of view is varied for MR-CT (top) and MR-PET (bottom)

		10 mm 10°			20 mm 20°			30 mm 30°		
	Measure	TF1	TF2	TF3	TF1	TF2	TF3	TF1	TF2	TF3
MR	$H(M, N)$	50	50	0	50	50	0	44	50	0
&	$I(M;N)$	42	50	50	16	50	50	3	50	48
CT	$Y(M;N)$	50	50	50	50	50	50	50	50	49

Studholme, Colin, Derek LG Hill, and David J. Hawkes. "An overlap invariant entropy measure of 3D medical image alignment." Pattern recognition 32.1 (1999): 71-86.

Contents

Motivation

Insight from maximum likelihood

A key point from the maximum likelihood interpretation is that a global joint intensity distribution is assumed:

$$p(r(s) = i, f \circ T_\alpha(s) = j; \alpha) = \beta_{ij} \geq 0, \quad \forall\, s \in \Omega_r^\alpha.$$

Then maximum mutual information finds the transformation that encourages a global functional relationship between the intensity pairs.

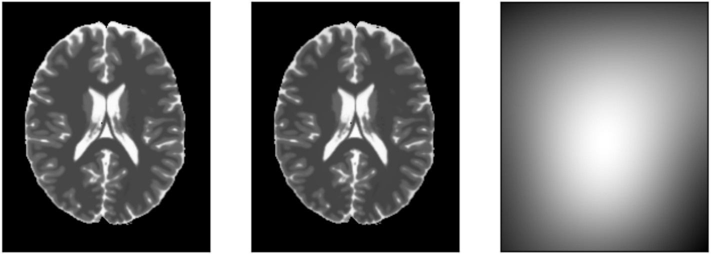
Remark

"This is only approximately true, e.g., when the images are distorted with a bias field or when structures with different intensities in one image have similar intensities in the other image."

Loeckx, Dirk, et al. "Nonrigid image registration using conditional mutual information." IEEE transactions on medical imaging 29.1 (2009): 19-29.

Examples

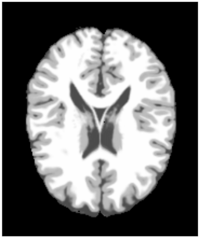
Example (Bias field)



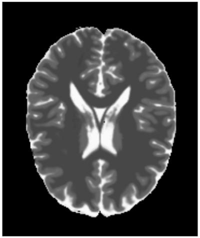
No INU   INU20%   Bias field

Examples

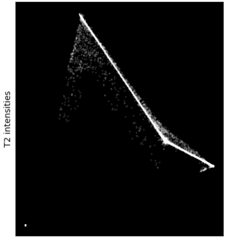
Joint histogram of clean images



T1 image



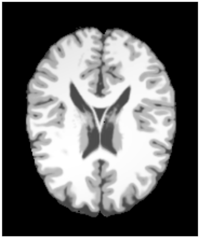
T2 image



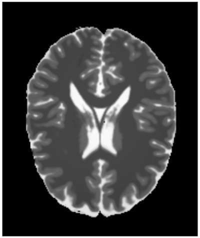
T1 intensities

Examples

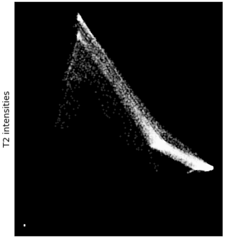
Joint histogram of corrupted images



T1 image



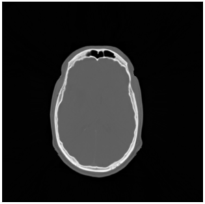
T2 image

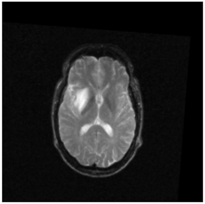


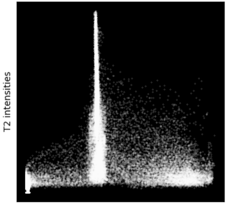
T1 intensities

Examples

Example (CT-MR)


CT

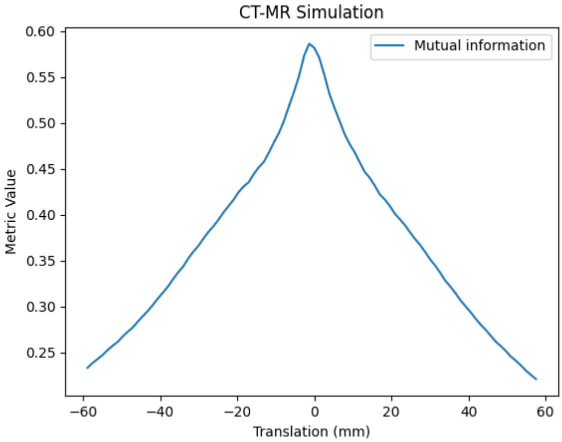

T2



CT intensities

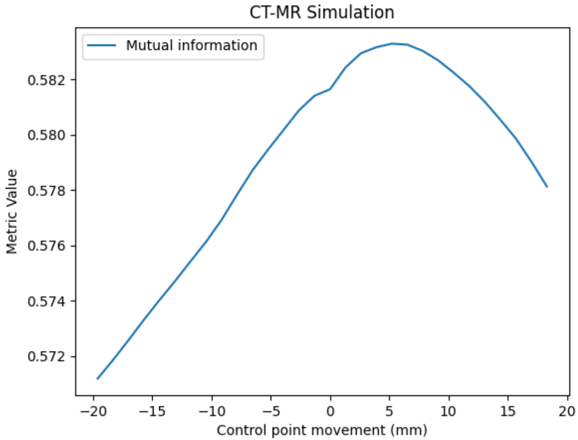
Simulation

Reponse of MI to translation



Simulation

Reponse of MI to FFD

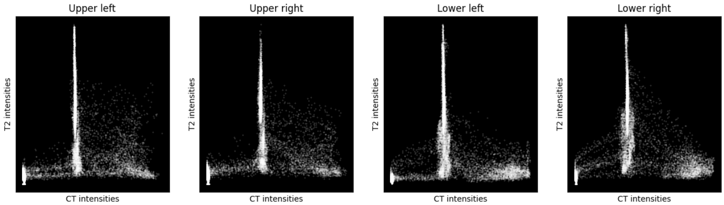


Motivation

Idea

Can the functional relationship be revealed in subregions of the images?

Example (CT-MR)



Literature

Related work

- Loeckx, Dirk, et al. "Nonrigid image registration using conditional mutual information." IEEE transactions on medical imaging 29.1 (2009): 19-29.
- Zhuang, Xiahai, et al. "A nonrigid registration framework using spatially encoded mutual information and free-form deformations." IEEE transactions on medical imaging 30.10 (2011): 1819-1828.
- Rivaz, Hassan, Zahra Karimaghaloo, and D. Louis Collins. "Self-similarity weighted mutual information: a new nonrigid image registration metric." Medical image analysis 18.2 (2014): 343-358.
- Woo, Jonghye, Maureen Stone, and Jerry L. Prince. "Multimodal registration via mutual information incorporating geometric and spatial context." IEEE Transactions on Image Processing 24.2 (2014): 757-769.

Conditional Mutual Information

Free-form deformation

$$T_\alpha(x,y,z) = \sum_{i,j,k} \boldsymbol{\rho}_{ijk} \cdot \beta^{n_x}_{\Delta_x}(x - u^x_i)\beta^{n_y}_{\Delta_y}(y - u^y_j)\beta^{n_z}_{\Delta_z}(z - u^z_k)$$

with $\Delta_{x,y,z}$ the mesh spacing, $n_{x,y,z}$ the B-spline degree, (u^x_i, u^y_j, u^z_k) the control point position, and $\boldsymbol{\rho}_{ijk}$ the control point displacement.

Conditional Mutual Information

Global mutual information

The Parzen window joint histogram for global MI is given by

$$\mathcal{H}_\alpha(\mu,\nu) = \sum_{s\in\Omega_r^\alpha} K_r\left(r(s)-\mu\right) K_f\left(f\circ T_\alpha(s)-\nu\right).$$

Conditional mutual information

To extend the joint histogram for conditional MI with spatial information:

$$\mathcal{H}_\alpha(\mu,\nu,\delta) = \sum_{s\in\Omega_r^\alpha} K_r\left(r(s)-\mu\right) K_f\left(f\circ T_\alpha(s)-\nu\right) K_s(s-\delta),$$

where δ is a spatial bin.

Conditional Mutual Information

Conditional mutual information (cMI)

The conditional MI is defined as

$$I(r, f \circ T_\alpha \mid s) \triangleq H(r \mid s) + H(f \circ T_\alpha \mid s) - H(r, f \circ T_\alpha \mid s)$$
$$= \sum_{\delta} p_\alpha(\delta) \sum_{\mu, \nu} p_\alpha(\mu, \nu \mid \delta) \ln \frac{p_\alpha(\mu, \nu \mid \delta)}{p_\alpha(\mu \mid \delta) p_\alpha(\nu \mid \delta)},$$

where

$$p_\alpha(\delta) = \frac{\sum_{\mu,\nu} \mathcal{H}_\alpha(\mu, \nu, \delta)}{\sum_{\mu,\nu,\delta'} \mathcal{H}_\alpha(\mu, \nu, \delta')}.$$

Discussion

Q&A

References

1. Pluim, Josien PW, JB Antoine Maintz, and Max A. Viergever. "Mutual-information-based registration of medical images: a survey." IEEE transactions on medical imaging 22.8 (2003): 986-1004.
2. Maes, Frederik, et al. "Multimodality image registration by maximization of mutual information." IEEE transactions on Medical Imaging 16.2 (1997): 187-198.
3. Viola, Paul, and William M. Wells III. "Alignment by maximization of mutual information." International journal of computer vision 24.2 (1997): 137-154.
4. Roche, Alexis, Gregoire Malandain, and Nicholas Ayache. "Unifying maximum likelihood approaches in medical image registration." International Journal of Imaging Systems and Technology 11.1 (2000): 71-80.
5. Studholme, Colin, Derek LG Hill, and David J. Hawkes. "An overlap invariant entropy measure of 3D medical image alignment." Pattern recognition 32.1 (1999): 71-86.
6. Loeckx, Dirk, et al. "Nonrigid image registration using conditional mutual information." IEEE transactions on medical imaging 29.1 (2009): 19-29.

Thank you!