MICCAI 2010 Tutorial

Intensity-based Deformable Registration

Similarity Measures

Christian Wachinger

Computer Aided Medical Procedures (CAMP), TU München, Germany
Quantify the similarity between images

• Do images contain the same object?

Image retrieval
Quantify the similarity between images

• How are images correctly aligned?

Image registration
Requirements on similarity measure

- Extremum for correctly aligned images
- Smooth, best convex
- Fast computation
- Differentiable
Overview

1. Standard similarity measures
2. Probabilistic framework
3. Pre-processing steps
4. Recent approaches
5. Linear vs. non-linear registration
PART I

Standard Similarity Measures

• SSD-SAD
• Cross Correlation
• Mutual information
• Derivatives
Difference Measures

\[ SSD = \frac{1}{N} \sum_i (x_i - y_i)^2 \]

Sum of Squared Differences

\[ SAD = \frac{1}{N} \sum_i |x_i - y_i| \]

Sum of Absolute Differences:
Less sensitive on large intensity differences than SSD

Volume X | Volume T(Y)
---|---

Volume X | Volume T(Y)
Limitations of SSD

- Illumination change affects similarity function

- Idea: normalization of images

\[ \tilde{X} = \frac{X - \mathbb{E}[X]}{\sigma(X)} \]
SSD on Normalized Images

\[ \mathbb{E}[(\tilde{X} - \tilde{Y})^2] \]

\[
\tilde{X} = \frac{X - \mathbb{E}[X]}{\sigma(X)} \\
\tilde{Y} = \frac{Y - \mathbb{E}[Y]}{\sigma(Y)}
\]
Normalized Cross Correlation (NCC)

\[ NCC = \frac{1}{N} \sum_{i} \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} \]

Normalized Cross Correlation:
Expresses the linear relationship between voxel intensities in the two volumes

\[ \bar{x} : \text{Mean} \]
\[ \sigma_x : \text{Standard deviation} \]
\[ N : \text{Number of pixels} \]
NCC - example

Sets of \((x,y)\) points, with the NCC of \(x\) and \(y\) for each set.

Source: wikipedia
Multi-Modal Registration

• More complex intensity relationship

• Approaches:
  – Simulate one modality from the other one
  – Apply sophisticated similarity measures
Information Theoretic Approach
Histogram calculation

Image

Histogram

Bins

4 3 3 6
1/4 3/16 3/16 3/8

counts
probs
Joint histogram calculation

Image X

Image Y

Overlap

\[
\begin{pmatrix}
  t_x = 1 \\
  t_y = 1
\end{pmatrix}
\]

Joint Histogram
Information Theoretic Approach

Registered

Not Registered

Source: W. Wells, MICCAI 2009
Joint Histogram

X and Y identical

\[ p_x(i) \]
\[ p_y(i) \]
\[ p_{xy}(i,j) \]

X and Y misaligned

\[ p_x(i) \]
\[ p_y(i) \]
\[ p_{xy}(i,j) \]
Joint Histogram

• Histogram for images from different Modalities

Source Image

Not Aligned

Aligned

Target Image

Joint Histogram

Joint Histogram

Intensities of Target $Y$

Intensities of Source $X$

NCC Optimum
$Y = a \cdot X + b$

SSD Optimum
$Y = X$
Information Theoretic Approach

- How to quantify the quality of alignment between MR and CT?
  - Measure the structure of the joint distribution
  - Shannon Entropy

Source: W. Wells, MICCAI 2009
Entropy

Shannon Entropy, developed in the 1940s
(communication theory)

\[ H = - \sum_{i} p_i \log p_i \]

uniform distribution
→ maximum entropy

any other distribution
→ less entropy
Maximize or minimize entropy?

Registered

Not Registered

CT

bone

white matter

fat

gray matter

CSF

CT

bone

white matter

fat

gray matter

CSF

air

air

any other distribution → less entropy

uniform distribution → maximum entropy

\[ p_i \rightarrow \text{entropy} \]
Mutual Information (MI)

\[ MI(X,Y) = H(X) + H(Y) - H(X,Y) \]

\[ = \sum_i \sum_j p_{xy}(i,j) \log \frac{p_{xy}(i,j)}{p_x(i)p_y(j)} \]

- Maximized if X and Y are perfectly aligned
- \( H(X) \) and \( H(Y) \) help to make the measure more robust
- Maximization of mutual information leads to minimization of joint entropy
Historical Note

• Minimum Entropy Registration

• Maximum Mutual Information Registration
Normalization of MI

• Problem: changing overlap affects MI

• Entropy Correlation Coefficient (ECC)

$\text{ECC}(X, Y) = \sqrt{2 - \frac{2 \cdot H(X, Y)}{H(X) + H(Y)}}$

• Normalized MI (NMI)

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

• Revisiting overlap invariance


NMI – change of overlap

• Two images

\[ T_2 \] has higher information content

• Overlap changes for deformable registration?

Improvements to MI

- Density estimation
  - Parzen window
  - Partial volume distribution
  - Uniform volume histogram
  - NP windows
- Spatial information
- Alternative entropy measures

- Tutorial at MICCAI 2009:  
  *Information theoretic similarity measures for image registration and segmentation: Maes, Wells, Pluim*
  
  [http://ubimon.doc.ic.ac.uk/MICCAI09/a1882.html](http://ubimon.doc.ic.ac.uk/MICCAI09/a1882.html)
Derivatives of Similarity Measures

- General form of derivative of similarity metrics

\[ \frac{\partial SM(X, Y(T_p))}{\partial p} = \frac{\partial SM(X, Y)}{\partial Y} \frac{\partial Y}{\partial T_p} \frac{\partial T_p}{\partial p} \]

- SSD:

\[ \nabla SSD(X, Y) = -2 \cdot (X - Y) \]

- MI

\[ \nabla MI(X, Y) = G_\Psi * \frac{1}{|\Omega|} \left( \frac{\partial^2 p(X, Y)}{p(X, Y)} - \frac{p'(Y)}{p(Y)} \right) \]

Derivatives of Similarity Measures

• General form of derivative of similarity metrics

\[
\frac{\partial SM(X, Y(T_p))}{\partial p} = \frac{\partial SM(X, Y)}{\partial Y} \frac{\partial Y}{\partial T_p} \quad \frac{\partial T_p}{\partial p}
\]
Derivatives of Similarity Measures

Images

\( \nabla \text{SSD} \)

\( \nabla Y \)

Update

\[
\begin{align*}
\frac{d}{dx} I_1 & \quad \frac{d}{dx} I_2 & \quad \frac{d}{dx} I_3 \\
\frac{d}{dy} I_1 & \quad \frac{d}{dy} I_2 & \quad \frac{d}{dy} I_3
\end{align*}
\]
Derivatives of Similarity Measures

Images

\( \nabla_{SSD} \)

\( \nabla Y \)

Update

\( \frac{\partial I}{\partial x} \)

\( \frac{\partial I}{\partial y} \)

\( d_{x1} \)

\( d_{x2} \)

\( d_{x3} \)

\( d_{y1} \)

\( d_{y2} \)

\( d_{y3} \)
Derivatives of Similarity Measures

Images

X

Y

−∇MI

∇Y

Update

\[
\begin{align*}
\frac{d}{dx} I_1 & \quad \frac{d}{dy} I_1 \\
\frac{d}{dx} I_2 & \quad \frac{d}{dy} I_2 \\
\frac{d}{dx} I_3 & \quad \frac{d}{dy} I_3
\end{align*}
\]
Overview

1. Standard similarity measures
2. Probabilistic framework
3. Pre-processing steps
4. Recent approaches
5. Linear vs. non-linear registration
PART II

Probabilistic Framework for Image Registration

- Viola, PhD thesis, 1995
- Roche et al., 2000
Probabilistic Framework for registration

• Maximum Likelihood Estimation (MLE)
  – Probability for the model \( m \) having the observations \( a \)

\[
\arg\max_m P(a = \{a_1, \ldots, a_n\}|m) = \arg\max_m \prod_{i=1}^n P(a_i|m)
\]

  – Log-likelihood function

\[
\log \mathcal{L}(m|a) = \sum_{i=1}^n \log P(a_i|m)
\]

• Formulate registration as likelihood maximization
MLE framework for registration

• Model

\[ X = f(Y(T)) + \varepsilon \]

• Probability function

\[ P(X|Y,T,\varepsilon,f) \]

• Log-likelihood function

\[
\log \mathcal{L}(T,\varepsilon,f) = \log P(X|Y,T,\varepsilon,f) \\
= \frac{1}{N} \sum_{v \in \Omega} \log P(X(v)|Y(T(v)),\varepsilon,f)
\]
\[
\log \mathcal{L}(T, \varepsilon, f) = \frac{1}{N} \sum_{v \in \Omega} \log P(X(v)|Y(T(v)), \varepsilon, f)
\]

- Stationary white Gaussian noise

\[
P(X|Y, T, f, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \cdot e^{-\frac{1}{2} \frac{[X-f(Y(T))]^2}{\sigma^2}}
\]

- Log-likelihood with Gaussian noise

\[
\log \mathcal{L}(T, \sigma, f) = -N \log \sqrt{2\pi\sigma} - \frac{1}{2\sigma^2} \frac{1}{N} \sum_{v \in \Omega} (X(v) - f(Y(T(v))))^2
\]

Assuming \( f \) to be the identity
\[
\log \mathcal{L}(T, \varepsilon, f) = \frac{1}{N} \sum_{v \in \Omega} \log P(X(v)|Y(T(v)), \varepsilon, f)
\]

- **Stationary white Gaussian noise**

\[
P(X|Y, T, f, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \cdot e^{-\frac{1}{2} \frac{[X-f(Y(T))]}{\sigma^2}^2}
\]

- **Log-likelihood with Gaussian noise**

\[
\log \mathcal{L}(T, \sigma) = -N \log \sqrt{2\pi\sigma} - \frac{1}{2\sigma^2} \frac{1}{N} \sum_{v \in \Omega} (X(v) - Y(T(v)))^2
\]

SSD
Intensity relationships

- Identity : SSD
- Affine : NCC
- Functional : Correlation Ratio
- Statistical : Mutual Information

Roche et al., *Unifying maximum likelihood approaches in medical image registration*, 2000
PART III

Pre-Processing Steps
Images → Registration Framework

Similarity Measure → Optimization
Images

Pre-processing

1. Image gradients
2. Entropy images
3. Phase
4. Multi-resolution
5. Attribute vectors

Registration Framework

Similarity Measure

Optimization
Image Gradients

• Normalized gradient fields

\[ n(X) := \begin{cases} \frac{\nabla X}{\|\nabla X\|}, & \text{if } \nabla X \neq 0 \\ 0, & \text{otherwise} \end{cases} \]

“two images are considered similar, if intensity changes occur at the same locations”

• Maximize square of cosine

\[ d(X, Y) = < n(X), n(Y) >^2 \]

Haber, E. and Modersitzki, J., *Intensity gradient based registration and fusion of multi-modal images*, MICCAI 2006
Image Gradients

Synthetic images

Gradient images
Image Gradients

Gradient of smoothed images

Gradient images
Images

Pre-processing

Registration Framework

1. Image gradients
2. Entropy images
3. Phase
4. Multi-resolution
5. Attribute vectors

Similarity Measure

Optimization
Entropy Images

Entropy Images

• Examples from the RIRE dataset
  
  http://www.insight-journal.org/rire/

Entropy Images

Synthetic images

Entropy images

Entropy Images

Entropy on smoothed images

Entropy images

Similarity Measure

Optimization

Images

Pre-processing

1. Image gradients
2. Entropy images
3. Phase
4. Multi-resolution
5. Attribute vectors

Registration Framework
Phase-base registration

• Fourier representation

\[ f(x) = \int_{0}^{\infty} A(\omega) \cdot \sin(\omega x + \phi(\omega))d\omega \]

\[ A(\omega) : \text{Amplitude} \]

\[ \phi(\omega) : \text{Phase} \]

M. Mellor and M. Brady. *Phase mutual information as a similarity measure for registration*. Medical Image Analysis, 2005.
Optical Flow with Phase

- Idea: replace the assumption of brightness constancy with phase constancy
- Error for changing noise and illumination

Images

Pre-processing

1. Image gradients
2. Entropy images
3. Phase
4. Multi-resolution
5. Attribute vectors

Registration Framework

Similarity Measure

Optimization
Multi-Resolution Registration

• Perform registration on multiple resolutions
  1. Smooth
  2. Downsample

• Advantages:
  – Speed: down-sampled images
  – Convergence: smoother cost func
Similarity Measure

Optimization

Registration Framework

Pre-processing

1. Image gradients
2. Entropy images
3. Phase
4. Multi-resolution
5. Attribute vectors

Images
Attribute Vectors

• Registration with attribute vectors
  \[ a = [ \text{gradient, intensity, GMI} ] \]

• GMI = geometric moment invariants


PART IV

Recent Approaches
Recent examples...

- mainly compare to MI as state of the art
- address the following problems of MI
  - intensity does not represent tissue (e.g. US-CT)
Recent examples...

- mainly compare to MI as state of the art
- address the following problems of MI
  - intensity does not represent tissue (e.g. US-CT)
  - intensity non-uniformity (bias)
Minimizing Residual Complexity

• Motivation: MI fails for these images

Myronenko A., Song X.: "Intensity-based Image Registration by Minimizing Residual Complexity", IEEE Trans. on Medical Imaging, 2010

Model

\[ X = Y(T) + S + \varepsilon \]
Minimizing Residual Complexity

- Compress the difference image

\[ \sum_{i=1}^{N} q_i \cdot (X - Y) \quad q_i: \ i\text{-th DCT basis function} \]

- Similarity measure

\[ RC(X, Y) = \sum_{i=1}^{N} \log \left( \frac{(q_i(X - Y))^2}{\alpha} + 1 \right) \]

- Related to entropy of difference image

\[ H(X - Y) \]

Myronenko A., Song X.: "Intensity-based Image Registration by Minimizing Residual Complexity", IEEE Trans. on Medical Imaging, 2010

Local and Global Statistics

\[ \text{NMI}_{l+g} = \frac{1}{|\Omega|} \sum_{x \in \Omega} (1 - \alpha_x) \cdot \text{NMI}_g + \alpha_x \cdot \text{NMI}_l(x) \]

\[ \alpha_x = \text{NMI}_l(x) - 1 \]

**Local and Global Statistics**

- Combining local and global densities
  \[ P(X, Y) = \alpha \cdot P_V(X, Y) + (1 - \alpha) \cdot P_{\overline{V}}(X, Y) \]

- \( \alpha \): manually, emphasize on local or global influence

---


PART V

Linear vs. Non-Linear Registration
Linear vs. Non-Linear Registration

\[
\frac{\partial SM(X, Y(T_p))}{\partial p} = \frac{\partial SM(X, Y)}{\partial Y} \frac{\partial Y}{\partial T_p} \frac{\partial T_p}{\partial p}
\]

- Deformable registration is more sensitive to errors from poor modeling
- Averaged out in linear registration
Linear vs. Non-Linear Registration

• Choice of similarity measure has influence on the optimization strategy

EXAMPLE: Gauß-Newton Optimization with SSD

\[ J_e^T J_e \Delta h = f \]
Linear vs. Non-Linear Registration

• Overlap invariance only for linear?
  – MI vs. NMI
  – Joint Entropy vs. MI

• Transition between linear and non-linear
  – Parameterization
  – Regularization
Summary

• A large variety of concepts for measuring the similarity is available in the literature
• Choose similarity measure that best fits the application
• Appropriate choice is more important for deformable registration
END